BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

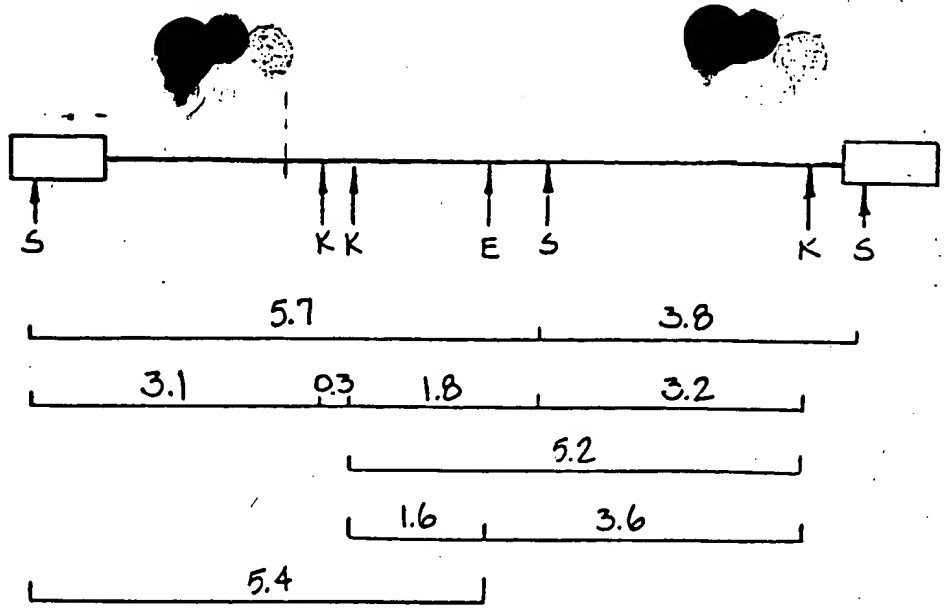


FIG. 1.

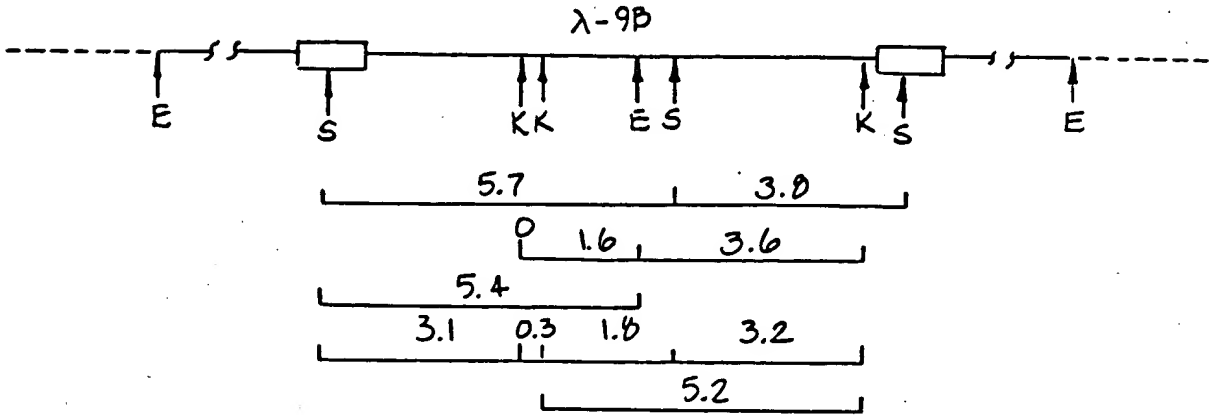


FIG. 2.

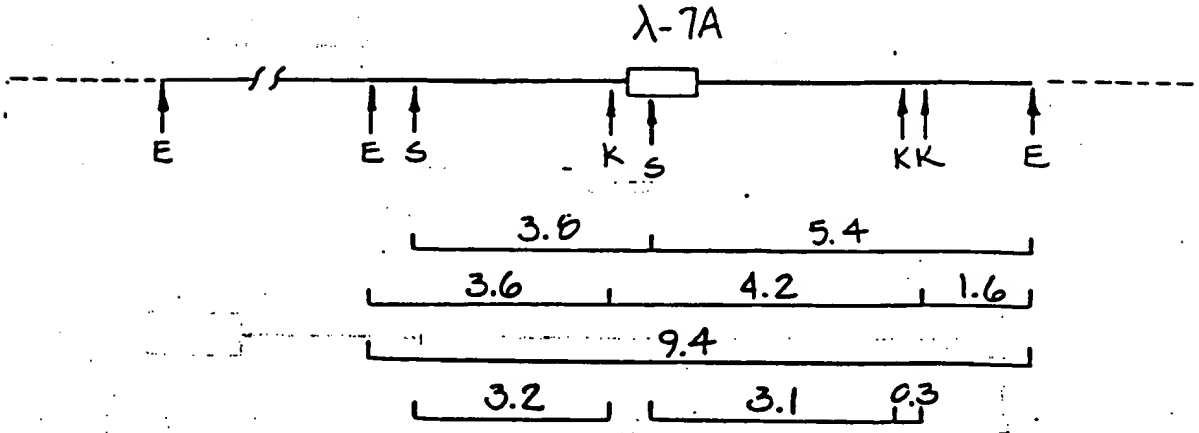
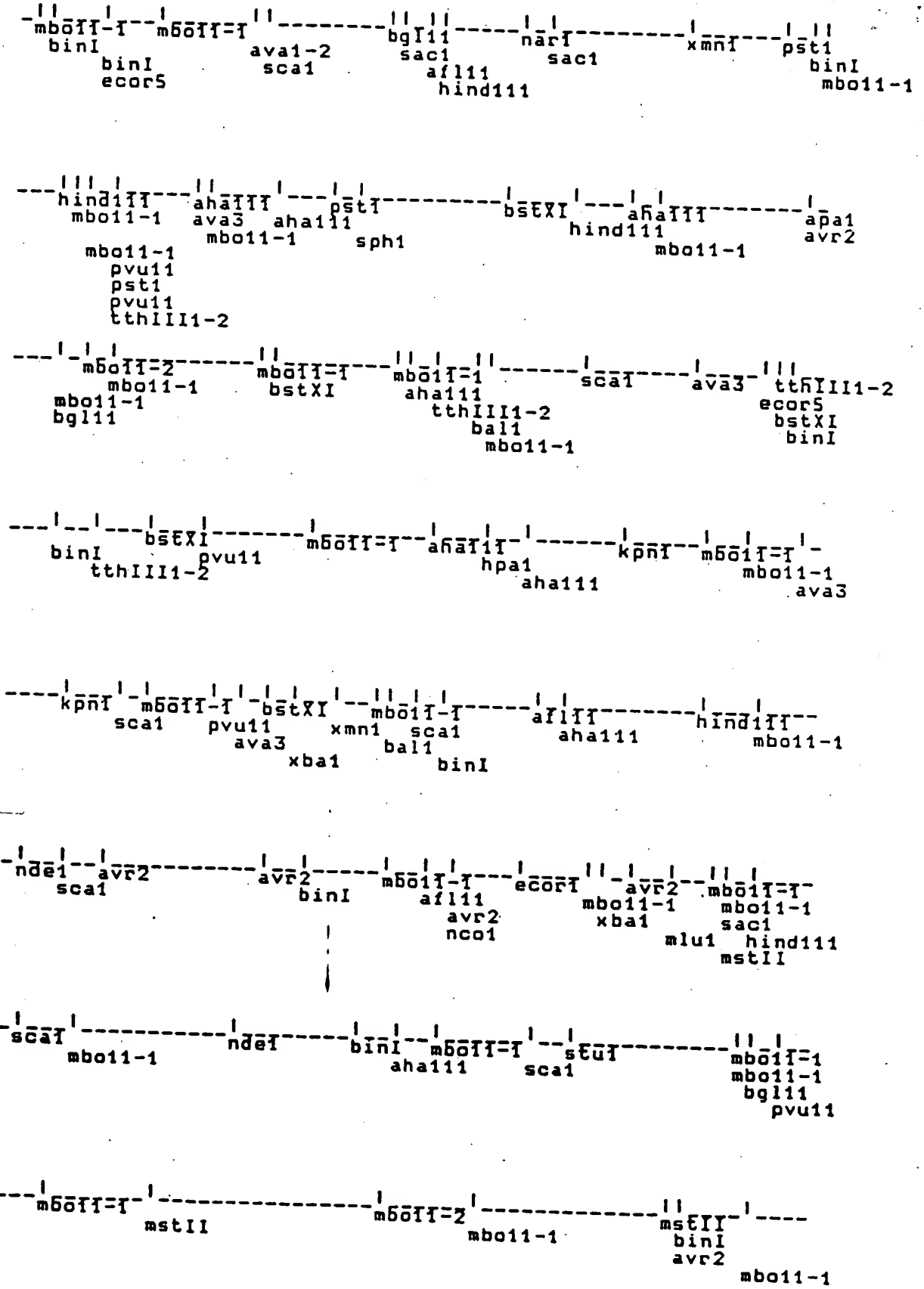
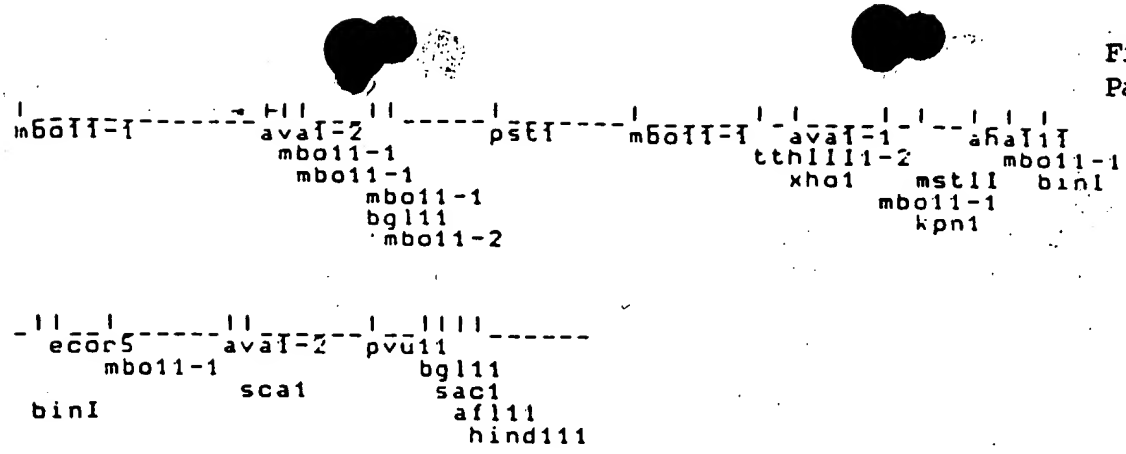


FIG. 3.

Argument Map in DNA Strand ssarv2
from the '/v/lib/6mers' file.
Translation shown at open reading frames.

FIGURE 4






```
1   CTGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACAC
    GACCTTCCCGATTAAACCAGGGTTTCTTCTGTTCTCTAGGAACTAGACACCTAGATGGTGTG

    26 mbo11, 50 binI,

63  ACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACT
    TGTTCCGATGAAGGGACTAACCGTCTTAATGTGTGGTCCCGGTCCCTAGTCTATAGGTGA

    107 binI, 113 ecor5,

123 GACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAA
    CTGGAAACCTACCACGAAGTTCGATCATGGTCAACTCGGTCTCTTCCATCTTCTCCGGTT

    172 mbo11,

183 TGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGA
    ACTTCCTCTCTTGTTGTCGAACAATGTGGGATACTCGGACGTACCCTACCTCCTGCGCCT

243 GAAAGAAGTGTTAGTGTGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGA
    CTTTCTTCACAATCACACCTCCAAACTGTCGTTTGATCGTAAAGTAGTGTACCGGGCTCT

    296 ava1,

303 GCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGC
    CGACGTAGGCCTCATGATGTTTCTGACGACTGTAGCTCGAAAGATGTTCCCTGAAAGGCG

    314 sca1,

363 TGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGC
    ACCCCTGAAAGGTCCCTCCGCACCGGACCCGCCCTGACCCCTCACCGCAGGGAGTCTACG

423 TGCATATAAGCAGACTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAG
    ACGTATATTCGTCTGACGAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTC

    474 bgl11,

483 CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
    GGACCCTCGAGAGACCGATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACGGAA.

    488 sac1, 518 afl11, 532 hind111,

543 GAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
    CTCACGAAGTTCATCACACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAGT

603 GACCCTTTTAGTCAGTGTGGAAAAATCTCTAGCAGTGGCGCCCGAACAGGGACGCGAAAG
    CTGGGAAAATCAGTCACACCTTTTTAGAGATCGTCACCGCGGGCTTGTCCCTGCGCTTTC

    639 nar1,

663 CGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACAG
    GCTTTCATCTTGGTCTCCTCGAGAGAGCTGCGTCCTGAGCCGAACGACTTCGCGCGTGTC

    680 sac1,

723 CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAG
    GTTCTCCGCTCCCCGCCGCTGACCACTCATGCGGTTAAAAACTGATCGCCTCCGATCTTC

                MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGlu
783 GAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA  GAG
    CTCTCTCTCTACCCACGCTCTCGCAGCCATAATTCGCCCCCTCTTAATCTATTTACCCTT
```

```
                LysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAla
 843            AAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAACATATAGTATGGGCA
                TTTTAAGCCAATTCCGGTCCCCCTTTCTTTTTTATATTCAATTTTGTATATCATACCCGT

                SerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCys
 903            AGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGCTGC
                TCGTCCCTCGATCTTGCTAAGCGTCAGTTAGGACCGGACAATCTTTGTAGTCTTCCGACG

                959 pst1,

                ArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSer
 963            AGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAAGAACTTAGATCA
                TCTGTTTATAACCCTGTCGATGTCGGTAGGGAAGTCTGTCCTAGTCTTCTTGAATCTAGT

                1002 binI, 1008 mbo11,

                LeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThr
1023            TTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACC
                AATATATTATGTCATCGTTGGGAGATAACACATGTAGTTTCCTATCTACATTTTCTGTGG

                LysGluAlaLeuGluLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGln
1083            AAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCAA
                TTCCTTCGAAATCTCTTCTATCTCCTTCTCGTTTTGTTTTCATTCTTTTTCCGTGTCGTT

                1087 hind111, 1097 mbo11, 1107 mbo11,                    p25

                AlaAlaAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleVal
1143            GCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAAAATTACCCTATAGTG
                CGTCGTCGACGTCGACCGTGTCCTTTGTCGTCGGTCCAGTCGGTTTTAATGGGATATCAC

                1147 pvu11, 1150 pst1, 1153 pvu11, 1156 tthIII1,

                GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
1203            CAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
                GTCTTGGATGTCCCCGTTTACCATGTAGTCCGGTATAGTGGATCTTGAAATTTACGTACC

                1250 aha111, 1255 ava3,

                ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu
1263            GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
                CATTTTCATCATCTTCTTTTCCGAAAGTCGGGTCTTCATTATGGGTACAAAAGTCGTAAT

                1275 mbo11,

                SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
1323            TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
                AGTCTTCCTCGGTGGGGTGTTCTAAATTTGTGGTACGATTTGTGTCACCCCCCTGTAGTT

                1346 aha111,

                AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal
1383            GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
                CGTCGGTACGTTTACAATTTTCTCTGATAGTTACTCCTTCGACGTCTTACCCTATCTCAC

                1423 pst1,

                HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
1443            CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC
                GTAGGTCACGTACGTCCCGGATAACGTGGTCCGGTTTACTCTCTTGGTTCCCCTTCACTG

                1451 sph1,

                IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro
1503            ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
                TATCGTCCTTGATGATCATGGGAAGTCCTTGTTTATCCTACCTACTGTTTATTAGGTGGA

                IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1563            ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA
                TAGGGTCATCCTCTTTAGATATTTTCTACCTATTAGGACCCTAATTTATTTTATCATTCT

                MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp
1623            ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT
                TACATATCGGGATGGTCGTAAGACCTGTATTCTGTTCCTGGTTTCCTTGGGAAATCTCTA

                1636 bstXI,

                TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1683            TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT
                ATACATCTGGCCAAGATATTTTGAGATTCTCGGCTTGTTCGAAGTGTCCTACATTTTTTA

                1720 hind111,
```

931154

```
                TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys
1743            TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA
                ACCTACTGTCTTTGGAACAACCAGGTTTTACGTTTGGGTCTAACATTCTGATAAAATTTT

                1796 aha111,

                AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1803            GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA
                CGTAACCCTGGTCGTCGATGTGATCTTCTTTACTACTGTCGTACAGTCCCTCACCCCCCT

                1827 mbo11,

                ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn
1863            CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC
                GGGCCGGTATTTCGTTCTCAAAACCGACTTCGGTACTCGGTTCATTGTTTAGGTCGATTG
                p18
                IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1923            ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT
                TATTACTACGTCTCTCCGTTAAAATCCTTGGTTTCTTTCTGACAATTCACAAAGTTAACA

                GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg
1983            GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA
                CCGTTTCTTCCCGTGTATCGGTTTTTAACGTCCCGGGGATCCTTTTTCCCGACAACCTCT

                2014 apa1, 2019 avr2,

                CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
2043            TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG
                ACACCTTCCCTTCCTGTGGTTTACTTTCTAACGTGACTCTCTGTCCGATTAAAAAATCCC

                2102 mbo11,

                LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro
2103            AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA
                TTCTAGACCGGAAGGATGTTCCCTTCCGGTCCCTTAAAAGAAGTCTCGTCTGGTCTCGGT

                2104 bgl11, 2141 mbo11,

                ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
2163            ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG
                TGTCGGGGTGGTCTTCTCTCGAAGTCCAAACCCCTCCTCTTTTGTTGAGGGAGAGTCTTC

                2175 mbo11,

                GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn
2223            CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC
                GTCCTCGGCTATCTGTTCCTTGACATAGGAAATTGAAGGGAGTCTAGTGAGAAACCGTTG

                AspProSerSerGlnOC
2283            GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGA
                CTGGGGAGCAGTGTTATTCCTATCCCCCCGTTGATTTCCTTCGAGATAATCTATGTCCT

                                MetAsnLeuProGlyLysTrpLysProLysMetIle
2342            GCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATA
                CGTCTACTATGTCATAATCTTCTTTACTTAAACGGTCCTTTTACCTTTGGTTTTTACTAT

                2360 mbo11, 2375 bstXI,

                GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCys
2402            GGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGT
                CCCCCTTAACCTCCAAAATAGTTTCATTCTGTCATGCTAGTCTATGGACATCTTTAGACA

                GlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg
2462            GGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGA
                CCTGTATTTCGATATCCATGTCATAATCATCCTGGATGTGGACAGTTGTATTAACCTTCT

                2517 mbo11,

                AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal
2522            AATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTA
                TTAGACAACTGAGTCTAACCAACATGAAATTTAAAGGGGTAATCAGGATAACTTTGACAT

                2548 aha111, 2577 tthIII,

                ProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu
2582            CCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAA
                GGTCATTTTAATTTCGGTCCTTACCTACCGGGTTTTCAATTCGTTACCGGTAACTGTCTT

                2627 ba11, 2639 mbo11,

                GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSer
2642            GAAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCA
                CTTTTTTATTTTCGTAATCATCTCTATACATGTCTTTACCTTTTCCTTCCCTTTTAAAGT
```

```
                LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer
2702            AAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGT
                TTTTAACCCGGACTTTTAGGTATGTTATGAGGTCATAAACGATATTTCTTTTTTCTGTCA

                2759 sca1,

                ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp
2762            ACTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGG
                TGATTTACCTCTTTTGATCATCTAAAGTCTCTTGAATTATTTTCTTGAGTTCTGAAGACC

                GluValGlnLeuGlyIleProHisProGlnGlyOC
2822            GAAGTTCAGTTAGGAATACCACACCCGCAGGGTTAAAAAAAGAAAAAATCAGTAACAGTA
                CTTCAAGTCAATCCTTATGGTGTGGGCGTCCCAATTTTTTTCTTTTTTAGTCATTGTCAT

2882            TTGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG
                AACCTACACCCACTACGTATGAAAAGTCAAGGGAATCTATTTCTGAAATCTTTCATATGAC

                2895 ava3,

                                        MetArgHisGlnGlyLeuAspIleSerThrMetTrp
2943            CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGG
                GTAAATGGTATGGATCATATTTGTTACTCTGTGGTCCCTAATCTATAGTCATGTTACACC

                2985 ecor5,

                LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeu
3003            CTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTA
                GACGGTGTCCCTACCTTTCCTAGTGGTCGTTATAAGGTTTCATCGTACTGTTTTTAGAAT

                3003 tthIII1, 3006 bstXI, 3021 binI,

                GluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr
3063            GAGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTAT
                CTCGGAAAATCTTTTGTCTTAGGTCTGTATCAATAGATAGTTATGTACCTACTAAACATA

                ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHis
3123            GTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCAT
                CATCCTAGACTGAATCTTTATCCCGTCGTATCTTGTTTTTATCTCCTTGACTCTGTCGTA

                3126 binI, 3171 tthIII1,

                LeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu
3183            CTGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTT
                GACAACTCCACCCCTAAATGGTGTGGTCTGTTTTTTGTAGTCTTTCTTGGAGGTAAGGAA

                3234 bstXI,

                TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGlu
3243            TGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAA
                ACCTACCCAATACTTGAGGTAGGACTATTTACCTGTCATGTCGGATATTACGACGGTCTT

                LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer
3303            AAAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGT
                TTTCTGTCGACCTGACAGTTACTGTATGTCTTCAATCACCCTTTTAACTTAACCCGTTCA

                3308 pvu11,

                GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAla
3363            CAGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCA
                GTCTAAATACGTCCCTAATTTCATTTCGTCAATACATTTGAGGAATCTCCTTGGTTTCGT

                LeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu
3423            CTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAG
                GATTGTCTTCATTATGGTGATTGTCTTCTTCGTCTCGATCTTGACCGTCTTTTGTCCCTC

                3447 mbo11,

                IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGlu
3483            ATTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAA
                TAAGATTTTCTTGGTCATGTACTTCATATAATACTGGGTAGTTTTCTGAATCATCGTCTT

                IleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn
3543            ATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAAT
                TATGTCTTCGTCCCCGTTCCGGTTACCTGTATAGTTTAAATAGTTCTCGGTAAATTTTTA

                3594 aha111,

                LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeu
3603            CTGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTA
                GACTTTTGTCCTTTCATACGTTCCTACTCCCCACGGGTGTGATTACTACATTTTGTCAAT

                3659 hpa1,
```

POL

```
                ThrGluAlaValGlnLysValSe. hrGluSerIleValIleTrpGlyLysIl.ProLys
3663            ACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAA
                TGTCTCCGTCACGTTTTTCATAGGTGTCTTTCGTATCATTATACCCCTTTCTAAGGATTT

                PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAla
3723            TTTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCT
                AAATTTGATGGGTATGTTTTCCTTTGTACCCTTCGTACCACCTACCTCATAACCGTTCGA

                3723 aha111,

                ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln
3783            ACCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAG
                TGGACCTAAGGACTCACCCTCAAACAGTTATGGGGAGGGAATCACTTTAATACCATGGTC

                3835 kpn1,

                LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg
3843            TTAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGG
                AATCTCTTTCTTGGGTATCATCCTCGTCTTTGAAAGATACATCTACCCCGTCGATTATCC

                GluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer
3903            GAGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCC
                CTCTGATTTAATCCTTTTCGTCCTATACAATGACTGTCTCCTTCTGTTTTTCAACAGAGG

                3943 mbo11,

                IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAsp
3963            ATAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGAT
                TATCGACTGTGTTGTTTAGTCTTCTGACTTAATGTTCGTTAAGTAGATCGAAACGTCCTA

                3983 mbo11,

                SerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla
4023            TCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCA
                AGCCCTAATCTTCATTTGTATCATTGTCTGAGTGTTATACGTAATCCTTAGTAAGTTCGT

                4060 ava3,

                GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLys
4083            CAACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAG
                GTTGGTCTATTCTCACTTAGTCTCAATCAGTCAGTTTATTATCTCGTCAATTATTTTTTC

                GluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal
4143            GAAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTA
                CTTTTCCAGATGGACCGTACCCATGGTCGTGTGTTTCCTTAACCTCCTTTACTTGTTCAT

                4163 kpn1,

                AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAla
4203            GATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCC
                CTATTTAATCAGTCACGACCTTAGTCCTTTCATGATAAAAACTTACCTTATCTATTCCGG

                4232 sca1,

                GlnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu
4263            CAAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTG
                GTTCTTCTTGTACTCTTTATAGTGTCATTAACCTCTCGTTACCGATCACTAAAATTGGAC

                4266 mbo11,

                ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGlu
4323            CCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAA
                GGTGGACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTCGATTTTCCTCTT

                4352 pvu11,

                AlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu
4383            GCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTA
                CGGTACGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTAACATGTGTAGAT

                4386 ava3, 4410 bstXI, 4439 xba1,

                GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluVal
4443            GAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTT
                CTTCCTTTTTAATAGGACCATCGTCAAGTACATCGGTCACCTATATATCTTCGTCTTCAA

                4497 xmn1,

                IleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp
4503            ATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGG
                TAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATAAAAGAGAATTTTAATCGTCCTTCTACC

                4555 mbo11, 4560 ba11,
```

```
          ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAla
4563      CCAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCC
          GGTCATTTTTGTTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATGATGCCAATTCCGG

          4605 sca1,

          AlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln
4623      GCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAA
          CGGACAACCACCCGTCCCTAGTTCGTCCTTAAACCGTAAGGGATGTTAGGGGTTTCAGTT

          4639 binI,

          GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGln
4683      GGAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAG
          CCTCATCATCTTAGATACTTATTACTTAATTTCTTTTAATATCCTGTCCATTCTCTAGTC

          AlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys
4743      GCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAA
          CGACTTGTGGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTAAAATTTTCTTTT

          4752 af111, 4791 aha111,

          GlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIle
4803      GGGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATA
          CCCCCCTAACCCCCTATGTCACGTCCCCTTTCTTATCATCTGTATTATCGTTGTCTGTAT

          GlnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg
4863      CAAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGG
          GTTTGATTTCTTGATGTTTTCGTTTAATGTTTTTAAGTTTTAAAAGCCCAAATAATGTCC

          AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAla
4923      GACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCA
          CTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGTTTCGAAGAGACCTTTCCACTTCCCCGT

          4956 hind111,

          ValValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle
4983      GTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATT
          CATCATTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTTCGTTTTTAGTAA

          5023 mbo11,

                MetGluAsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIle
          ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAsp
5043      AGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGAT
          TCCCTAATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCTGTCCTACTCCTA

          ArgTreTrpLysSerLeuValLysHisHisMetTyrIleSerLysLysAlaLysGlyTrp
          AM
5103      TAGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG
          ATCTTGTACCTTTTCAAATCATTTTGTGGTATACATATAAAGTTTCTTTCGATTTCCTACC

          5131 nde1,

             PheTyrArgHisHisTyrGluSerThrHisProArgValSerSerGluValHisIle
5163         TTTTATAGACATCACTATGAAAGTACTCATCCAAGAGTAAGTTCAGAAGTACACATC
             AAAATATCTGTAGTGATACTTTCATGAGTAGGTTCTCATTCAAGTCTTCATGTGTAG

          5185 sca1,

          ProLeuGlyAspAlaLysLeuValIleThrThrTyrTrpGlyLeuHisThrGlyGluArg
5221      CCCCTAGGGGATGCTAAATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGA
          GGGGATCCCCTACGATTTAACCATTATTGTTGTATAACCCCAGACGTATGTCCTCTTTCT

          5223 avr2,

          GluTrpHisLeuGlyGlnGlyValAlaIleGluTrpArgLysLysLysTyrSerThrGln
5281      GAATGGCATTTGGGCCAGGGAGTCGCCATAGAATGGAGGAAAAAGAAATATAGCACACAA
          CTTACCGTAAACCCGGTCCCTCAGCGGTATCTTACCTCCTTTTTCTTTATATCGTGTGTT

          ValAspProGlyLeuAlaAspGlnLeuIleHisLeuHisTyrPheAspCysPheSerGlu
5341      GTAGACCCTGGCCTAGCAGACCAACTAATTCATCTGCATTATTTTGATTGTTTTTCAGAA
          CATCTGGGACCGGATCGTCTGGTTGATTAAGTAGACGTAATAAAACTAACAAAAAGTCTT

          SerAlaIleLysAsnAlaIleLeuGlyTyrArgValSerProArgCysGluTyrGlnAla
5401      TCTGCTATAAAAAATGCCATATTAGGATATAGAGTTAGTCCTAGGTGTGAATATCAAGCA
          AGACGATATTTTTTACGGTATAATCCTATATCTCAATCAGGATCCACACTTATAGTTCGT

          5440 avr2,

          GlyHisAsnLysValGlySerLeuGlnTyrLeuAlaLeuAlaAlaLeuIleThrProLys
5461      GGACATAACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAA
          CCTGTATTGTTCCATCCTAGAGATGTTATGAACCGTGATCGTCGTAATTATTGTGGTTTT

          5476 binI,
```

```
                LysThrLysProProLeuProSerValLysLysLeuThrGluAspArgTrpAsnLysPro
5521  AAGACAAAGCCACCTTTGCCTAGTGTTAAGAAACTGACAGAGGATAGATGGAACAAGCCC
      TTCTGTTTCGGTGGAAACGGATCACAATTCTTTGACTGTCTCCTATCTACCTTGTTCGGG

                GlnLysThrLysGlyHisArgGlySerHisThrMetAsnGlyHisAM
5581  CAGAAGACCAAGGGCCACAGAGGGAGCCATACAATGAATGGACACTAGAGCTTTTAGAGG
      GTCTTCTGGTTCCCGGTGTCTCCCTCGGTATGTTACTTACCTGTGATCTCGAAAATCTCC

      5583 mbo11,

5641  AGCTTAAGAGAGAAGCTGTTAGACATTTTCCTAGGCCATGGCTCCATAGCTTAGGACAAT
      TCGAATTCTCTCTTCGACAATCTGTAAAAGGATCCGGTACCGAGGTATCGAATCCTGTTA

      5643 af111, 5670 avr2, 5676 nco1,

5701  ATATCTATGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGC
      TATAGATACTTTGAATACCCCTATGAACCCGTCCTCACCTTCGGTATTATTCTTAAGACG

      5752 ecor1,

5761  AACAACTGCTGTTTATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTATTC
      TTGTTGACGACAAATAAGTAAAGTCTTAACCCACAGTTGTATCGTCTTATCCGTAATAAG

5821  AACAGAGGAGAGCAAGAAGAAATGGAGCCAGTAGATCCTAATCTAGAGCCCTGGAAGCAT
      TTGTCTCCTCTCGTTCTTCTTTACCTCGGTCATCTAGGATTAGATCTCGGGACCTTCGTA

      5836 mbo11, 5862 xba1,

5881  CCAGGAAGTCAGCCTAGGACTGCTTGTAACAATTGCTATTGTAAAAAGTGTTGCTTTCAT
      GGTCCTTCAGTCGGATCCTGACGAACATTGTTAACGATAACATTTTTCACAACGAAAGTA

      5893 avr2,

5941  TGCTACGCGTGTTTCACAAGAAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGA
      ACGATGCGCACAAAGTGTTCTTTTCCGAATCCGTAGAGGATACCGTCCTTCTTCGCCTCT

      5945 mlu1, 5988 mbo11,

6001  CAGCGACGAAGAGCTCCTCAGGACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAA
      GTCGCTGCTTCTCGAGGAGTCCTGTCAGTCTGAGTAGTTCGAAGAGATAGTTTCGTCATT

      6008 mbo11, 6011 sac1, 6016 mstII, 6038 hind111,

6061  GTAGTAAATGTAATGCAATCTTTACAAATATTAGCAATAGTATCATTAGTAGTAGTAGCA
      CATCATTTACATTACGTTAGAAATGTTTATAATCGTTATCATAGTAATCATCATCATCGT

6121  ATAATAGCAATAGTTGTGTGGACCATAGTACTCATAGAATATAGGAAAATATTAAGACAA
      TATTATCGTTATCAACACACCTGGTATCATGAGTATCTTATATCCTTTTATAATTCTGTT

      6147 sca1,

                                                              MetLys   ENV
6181  AGAAAATAGACAGATTAATTGATAGAATAAGAGAAAAAGCAGAAGACAGTGGCAATGAAA
      TCTTTTATCTGTCTAATTAACTATCTTATTCTCTTTTTCGTCTTCTGTCACCGTTACTTT

      6222 mbo11,

                ValLysGlyThrArgArgAsnTyrGlnHisLeuTrpArgTrpGlyThrLeuLeuLeuGly
6241  GTGAAGGGGACCAGGAGGAATTATCAGCACTTGTGGAGATGGGGCACCTTGCTCCTTGGG
      CACTTCCCCTGGTCCTCCTTAATAGTCGTGAACACCTCTACCCCGTGGAACGAGGAACCC

                MetLeuMetIleCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValPro
6301  ATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCT
      TACAACTACTAGACATCACGATGTCTTTTTAACACCCAGTGTCAAATAATACCTCATGGA

                ValTrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThr
6361  GTGTGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACA
      CACACCTTTCTTCGTTGATGGTGAGATAAAACACGTAGTCTACGATCTCGTATACTATGT

      6410 nde1,

                GluValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGlu
6421  GAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAA
      CTCCATGTATTACAAACCCGGTGTGTACGGACACATGGGTGTCTGGGGTTGGGTGTTCTT

                ValValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGln
6481  GTAGTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAG
      CATCATAACCCTTTACACTGTCTTTTAAAATTGTACACCTTTTTATTGTACCATCTTGTC

                MetGlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThr
6541  ATGCAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACC
      TACGTCCTCCTATATTAGTCAAATACCCTAGTTTCGGATTTCGGTACACATTTTAATTGG

      6567 bin1,
```

```
                ProLeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSer
5601            CCACTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGT
                GGTGAGACACAATGAAATTTAACGTGACTAAACCCCTTCCGATGATTATGGTTATCATCA

                6615 aha111,

                AsnTrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSer
5661            AATTGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGC
                TTAACCTTTCTTCTTTATTTTCCTCTTTATTTTTTGACGAGAAAGTTATAGTGGTGTTCG

                6670 mbo11,

                IleArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIle
5721            ATAAGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATA
                TATTCTCTATTCTAAGTCTTTCTTTTACGTGAAAAAGCATTGGAACTACATCATGGTTAT

                AspAsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSer
5781            GATAATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCA
                CTATTACGATCATGATGATGGTTGATATGGTTGATATCCAACTATGTAACATTGTCTAGT

                6790 sca1,

                ValIleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThr
5841            GTCATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACC
                CAGTAATGTGTCCGGACAGGTTTCCATAGTAAACTCGGTTAAGGGTATGTAATAACATGG

                6851 stu1,

                ProAlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCys
5901            CCGGCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGT
                GGCCGACCAAAACGCTAAGATTTCACATTATTATTTTGCAAGTTACCTTTTCCTGGTACA

                ThrAsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeu
5961            ACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTG
                TGTTTACAGTCGTGTCATGTTACATGTGTACCTTAATCCGGTTATCACAGTTGAGTTGAC

                LeuLeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsn
7021            CTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAAC
                GACAATTTACCGTCAGATCGTCTTCTTCTCCATCATTAATCTAGACTGTTAAAGTGCTTG

                7042 mbo11, 7045 mbo11, 7060 bgl11,

                AsnAlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgPro
7081            AATGCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCC
                TTACGATTTTGGTATTATCATGTCGACTTACTTAGACATCGTTAATTGACATGTTCTGGG

                7102 pvu11,

                AsnAsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGly
7141            AACAACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGA
                TTGTTGTTATGTTCTTTTTCATAGATATATCCTGGTCCCTCTCGTAAAGTATGTTGTCCT

                7199 mbo11,

                ArgIleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsn
7201            AGAATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAAC
                TCTTATTATCCTCTATATTCTTTTCGTGTAACATTGTAATCATCTCGTGTTACCTTATTG

                ThrLeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleVal
7261            ACTTTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTC
                TGAAATCTTGTCTATCAATTTTTTAATTCTCTTGTCAAACCCTTATTATTTTGTTATCAG

                PheAsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGly
7321            TTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGG
                AAATTAGTTAGGAGTCCTCCCCTGGGTCTTTAACATTACGTGTCAAAATTAACATCTCCC

                7331 mstII,

                GluPhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThr
7381            GAATTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACT
                CTTAAAAAAGATGACATTATGTTGTGTTGACAAATTATTATGTACCTCCAATTTAGTGTGA

                GluGlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsn
7441            GAAGGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAAC
                CTTCCTTGATTTCCTTTACTGTGTTAGTATGAGGGTACATCTTATTTTGTTTAATATTTG

                MetTrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCys
7501            ATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGT
                TACACCGTCCTTCATCCTTTTCGTTACATACGGGGAGGGTAACCTCCTGTTTAATCAACA

                SerSerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAsp
7561            TCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGAC
                AGTAGTTTATAATGTCCCGACGATAATTGTTCTCTACCACCATGTTTACATTGATTACTG
```

```
                ThrGluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyr
7621            ACCGAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATAT
                TGGCTCCAGAAGTCTGGACCTCCTCCTCTATACTCCCTGTTAACCTCTTCACTTAATATA

                7628 mbo11,

                LysTyrLysValIleLysIleGluProLeuGlyIleAlaProThrLysAlaLysArgArg
7681            AAATATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACCAAGGCAAAGAGAAGA
                TTTATATTTCATTATTTTTAACTTGGTAATCCTTATCGTGGGTGGTTCCGTTTCTCTTCT

                7736 mbo11,

                ValValGlnArgGluLysArgAlaValGlyIleValGlyAlaMetPheLeuGlyPheLeu
7741            GTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGTAGGAGCTATGTTCCTTGGGTTCTTG
                CACCACGTCTCTCTTTTTTCTCGTCACCCTTATCATCCTCGATACAAGGAACCCAAGAAC

                GlyAlaAlaGlySerThrMetGlyAlaValSerLeuThrLeuThrValGlnAlaArgGln
7801            GGAGCAGCAGGAAGCACTATGGGCGCAGTGTCATTGACGCTGACGGTACAGGCCAGACAA
                CCTCGTCGTCCTTCGTGATACCCGCGTCACAGTAACTGCGACTGCCATGTCCGGTCTGTT

                LeuLeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGln
7861            TTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAA
                AATAACAGACCATATCACGTTGTCGTCTTGTTAAACGACTCCCGATAACTCCGCGTTGTT

                HisLeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaVal
7921            CATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTG
                GTAGACAACGTTGAGTGTCAGACCCCGTAGTTCGTCGAGGTCCGTTCTCAGGACCGACAC

                GluArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIle
7981            GAAAGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATT
                CTTTCTATGGATTCCCTAGTTGTCGAGGATCCCTAAACCCCAACGAGACCTTTTGAGTAA

                7989 mstII, 7995 binI, 8007 avr2,

                CysThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrp
8041            TGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAAGACATTTGG
                ACGTGGTGACGACACGGAACCTTACGATCAACCTCATTATTTAGAGACCTTCTGTAAACC

                8089 mbo11,

                AspAsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyr
8101            GATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAACACAATATAC
                CTATTGTACTGGACCTACGTCACCCTTTCTCTTTAACTGTTAATGTGTTTGTGTTATATG

                ThrLeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeu
8161            ACCTTACTTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTG
                TGGAATGAACTTCTTAGCGTCTTGGTTGTTCTTTTCTTACTTGTTCTTAATAATCTTAAC

                8170 mbo11,

                AspLysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpLeuTrpTyrIleLys
8221            GATAAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAAG
                CTATTCACCCGTTCAAACACCTTAACCAAATCGTATTGTTTGACCGACACCATATATTTC

                IlePheIleMetIleValGlyGlyLeuValGlyLeuArgIleValPheAlaValLeuSer
8281            ATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTCT
                TATAAGTATTACTATCATCCTCCGAACCATCCAAATTCTTATCAAAAACGACACGAAAGA

                IleValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrArgLeuProVal
8341            ATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGTCATTTCAGACCCGCCTCCCAGTC
                TATCACTTATCTCAATCCGTCCCTATGAGTGGTAACAGTAAAGTCTGGGCGGAGGGTCAG

                8400 ava1,

                ProArgGlyProAspArgProAspGlyIleGluGluGluGlyGlyGluArgAspArgAsp
8401            CCGAGGGGACCCGACAGGCCCGACGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGAC
                GGCTCCCCTGGGCTGTCCGGGCTGCCTTAGCTTCTTCTTCCACCTCTCTCTCTGTCTCTG

                8431 mbo11, 8434 mbo11,

                ArgSerValArgLeuValAspGlyPheLeuAlaLeuIleTrpGluAspLeuArgSerLeu
8461            AGATCCGTTCGATTAGTGGATGGATTCTTAGCACTTATCTGGGAAGATCTGCGGAGCCTG
                TCTAGGCAAGCTAATCACCTACCTAAGAATCGTGAATAGACCCTTCTAGACGCCTCGGAC

                8503 mbo11, 8505 bgl11,

                CysLeuPheSerTyrArgArgLeuArgAspLeuLeuLeuIleAlaAlaArgThrValGlu
8521            TGCCTCTTCAGCTACCGCCGCTTGAGAGACTTACTCTTGATTGCAGCGAGGACTGTGGAA
                ACGGAGAAGTCGATGGCGGCGAACTCTCTGAATGAGAACTAACGTCGCTCCTGACACCTT

                8525 mbo11,
```

```
                IleLeuGlyHisArgGlyTrpG..AlaLeuLysTyrTrpTrpSerLeuLeuG.nTyrTrp
8581            ATTCTGGGGCACAGGGGGTGGGAAGCCCTCAAATATTGGTGGAGTCTCCTGCAGTATTGG
                TAAGACCCCGTGTCCCCCACCCTTCGGGAGTTTATAACCACCTCAGAGGACGTCATAACC

                8629 pst1,

                IleGlnGluLeuLysAsnSerAlaValSerTrpLeuAsnAlaThrAlaIleAlaValThr
8641            ATTCAGGAACTAAAGAATAGTGCTGTTAGCTGGCTCAACGCCACAGCTATAGCAGTAACT
                TAAGTCCTTGATTTCTTATCACGACAATCGACCGAGTTGCGGTGTCGATATCGTCATTGA

                GluGlyThrAspArgValIleGluValAlaGlnArgAlaTyrArgAlaIleLeuHisIle
8701            GAGGGGACAGATAGGGTTATAGAAGTAGCACAAAGAGCTTATAGAGCTATTCTCCACATA
                CTCCCCTGTCTATCCCAATATCTTCATCGTGTTTCTCGAATATCTCGATAAGAGGTGTAT

                HisArgArgIleArgGlnGlyLeuGluArgLeuLeuLeuOC  MetGlyGlyLysTrpSer
8761            CATAGAAGAATTAGACAGGGCTTGGAAAGGCTTTTGCTATAAGATGGGTGGCAAGTGGTCA
                GTATCTTCTTAATCTGTCCCGAACCTTTCCGAAAACGATATTCTACCCACCGTTCACCAGT

                8765 mbo11,

                LysArgSerMetGlyGlyTrpSerAlaIleArgGluArgMetArgArgAlaGluProArg
8822            AAACGTAGTATGGGTGGATGGTCTGCTATAAGGGAAAGAATGAGACGAGCTGAGCCACGA
                TTTGCATCATACCCACCTACCAGACGATATTCCCTTTCTTACTCTGCTCGACTCGGTGCT

                AlaGluProAlaAlaAspGlyValGlyAlaValSerArgAspLeuGluLysHisGlyAla
8882            GCTGAGCCAGCAGCAGATGGGGTGGGAGCAGTATCTCGAGACCTGGAAAAACATGGAGCA
                CGACTCGGTCGTCGTCTACCCCACCCTCGTCATAGAGCTCTGGACCTTTTTTGTACCTCGT

                8883 tthIII1, 8916 ava1 xho1,

                IleThrSerSerAsnThrAlaAlaThrAsnAlaAspCysAlaTrpLeuGluAlaGlnGlu
8942            ATCACAAGTAGCAATACAGCAGCTACTAATGCTGATTGTGCCTGGCTAGAAGCACAAGAG
                TAGTGTTCATCGTTATGTCGTCGATGATTACGACTAACACGGACCGATCTTCGTGTTCTC

                GluGluGluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLys
9002            GAGGAAGAGGTGGGTTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTACAAG
                CTCCTTCTCCACCCAAAAGGTCAGTCTGGAGTCCATGGAAATTCTGGTTACTGAATGTTC

                9005 mbo11, 9029 mstII, 9034 kpn1,

                AlaAlaLeuAspIleSerHisPheLeuLysGluLysGlyGlyLeuGluGlyLeuIleTrp
9062            GCAGCTTTAGATATTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTTGG
                CGTCGAAATCTATAATCGGTGAAAAATTTTCTTTTCCCCCCTGACCTTCCCGATTAAACC

                9085 aha111,

                SerGlnArgArgGlnGluIleLeuAspLeuTrpIleTyrHisThrGlnGlyTyrPhePro
9122            TCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCT
                AGGGTTTCTTCTGTTCTCTAGGAACTAGACACCTAGATGGTGTGTGTTCCGATGAAGGGA

                9129 mbo11, 9153 binI,

                AspTrpGlnAsnTyrThrProGlyProGlyIleArgTyrProLeuThrPheGlyTrpCys
9182            GATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
                CTAACCGTCTTAATGTGTGGTCCCGGTCCCTAGTCTATAGGTGACTGGAAACCTACCACG

                9210 binI, 9216 ecor5,

                PheLysLeuValProValGluProGluLysValGluGluAlaAsnGluGlyGluAsnAsn
9242            TTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAAC
                AAGTTCGATCATGGTCAACTCGGTCTCTTCCATCTTCTCCGGTTACTTCCTCTCTTGTTG

                9275 mbo11,

                SerLeuLeuHisProMetSerLeuHisGlyMetGluAspAlaGluLysGluValLeuVal
9302            AGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGAGAAAGAAGTGTTAGTG
                TCGAACAATGTGGGATACTCGGACGTACCCTACCTCCTGCGCCTCTTTCTTCACAATCAC

                TrpArgPheAspSerLysLeuAlaPheHisHisMetAlaArgGluLeuHisProGluTyr
9362            TGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTAC
                ACCTCCAAACTGTCGTTTGATCGTAAAGTAGTGTACCGGGCTCTCGACGTAGGCCTCATG

                9399 ava1, 9417 sca1,

                TyrLysAspCysOP
9422            TACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGG
                ATGTTTCTGACGACTGTAGCTCGAAAGATGTTCCCTGAAAGGCGACCCCTGAAAGGTCCC

9482            AGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGCTGCATATAAGCAGCTG
                TCCGCACCGGACCCGCCCTGACCCCTCACCGCAGGGAGTCTACGACGTATATTCGTCGAC

                9536 pvu11,
```

```
9542  CTTTTTGCCTGTACTGGGTCTC...GGTTAGACCAGATCTGAGCCTGGGAGC..TCTGGC
      GAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTCGGACCCTCGAGAGACCG

      9576 bgl11, 9590 sac1,

9602  TAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTG
      ATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACGGAACTCACGAAGTTCATCAC

      9620 af111, 9634 hind111,

9662  TGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTG
      ACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAGTCTGGGAAAATCAGTCAC

9722  TGGAAAAATCTCTAGCAG
      ACCTTTTTAGAGATCGTC
```

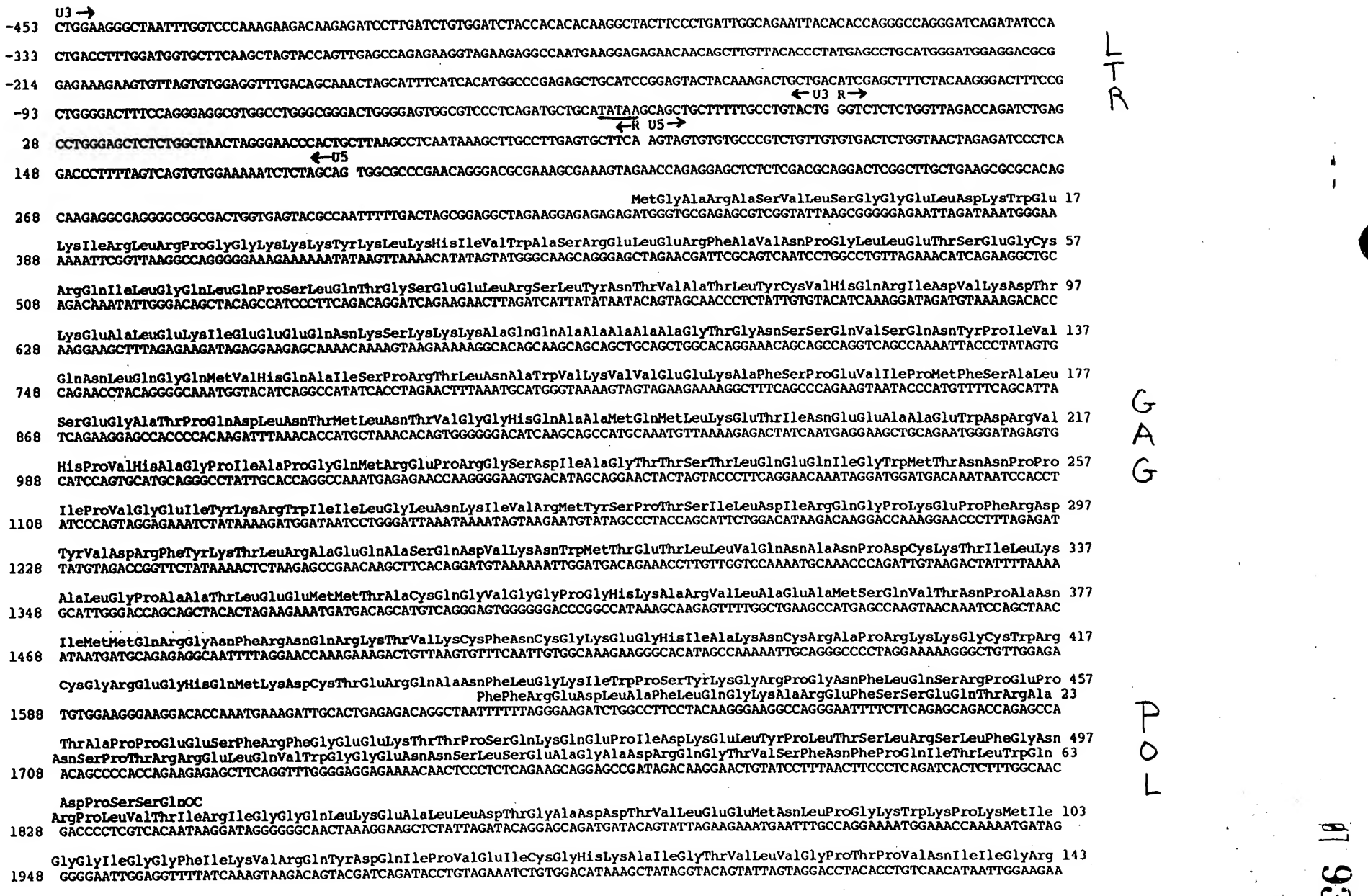


FIGURE 5
1 of 4

```
                AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 183
2068  ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

                GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
2188  AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

                ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
2308  CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGGAAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

                LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
2428  TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

                LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
2548  TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

                ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
2668  TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATCTGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

                TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 423
2788  GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

                GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463
2908  AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA

                IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503
3028  TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

                LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543
3148  TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTAACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAAT

                PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 583
3268  TTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCTACCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAGT

                LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer 623
3388  TAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCCA

                IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 663
3508  TAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCAC

                GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 703
3628  AACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAG

                AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAlaGlnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu 743
3748  ATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCCCAAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGC

                ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu 783
3868  CACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTAG

                GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp 823
3988  AAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGGC

                ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln 863
4108  CAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAAG

                GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys 903
4228  GAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAG

                GlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg 943
4348  GGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGG

                AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle 983
4468  ACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATTA

                ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAspAM
4588  GGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG
```

FIGURE 5
2 of 4

POL

EP 931154

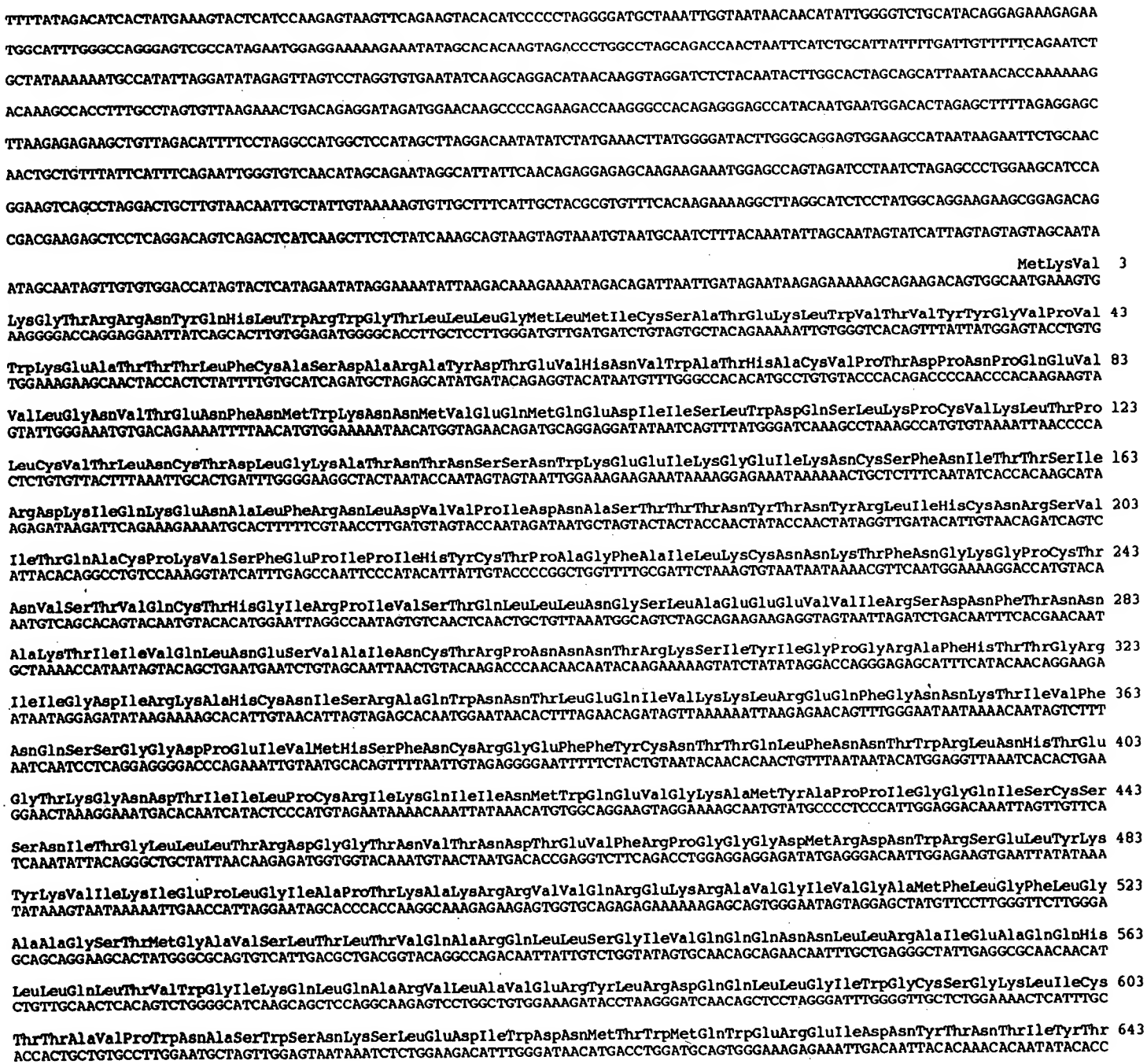


FIGURE 5
3 of 4

ENV

931154

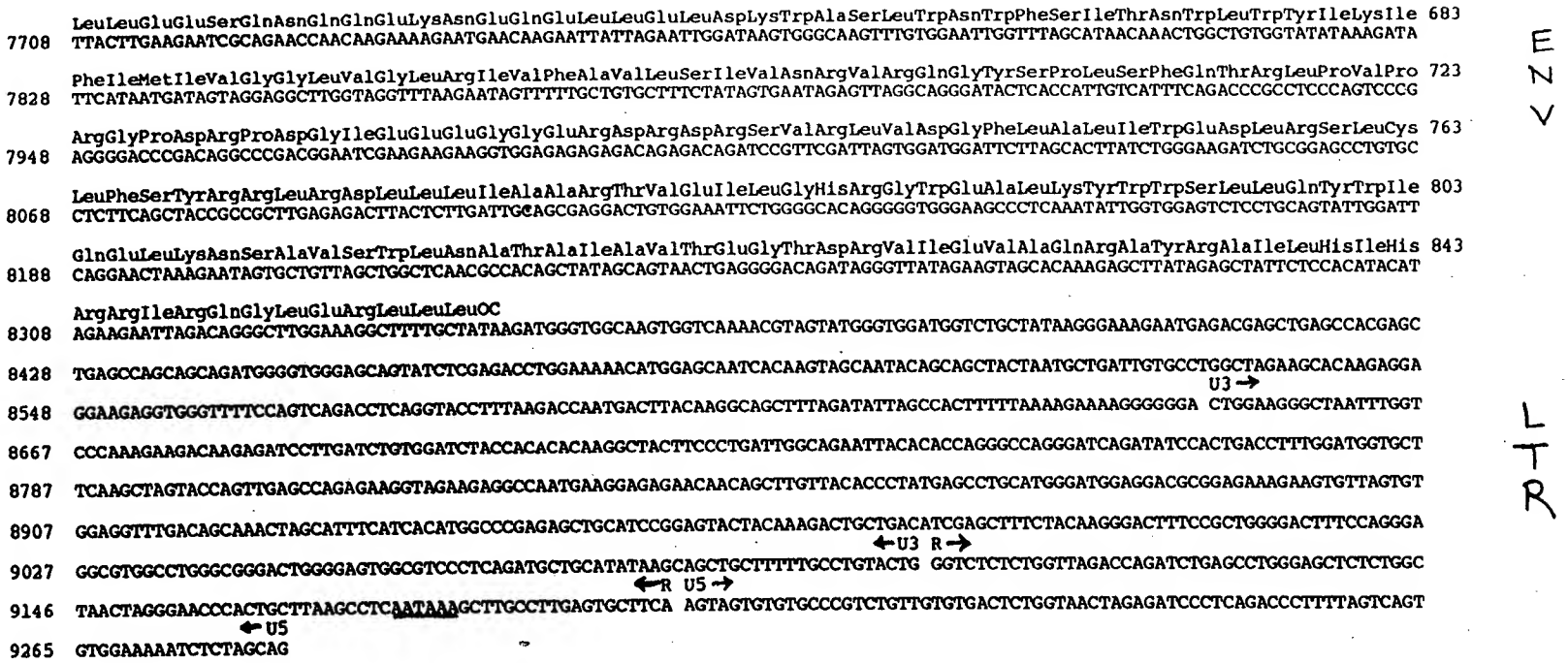


FIGURE 5
4 of 4

ENV

LTR

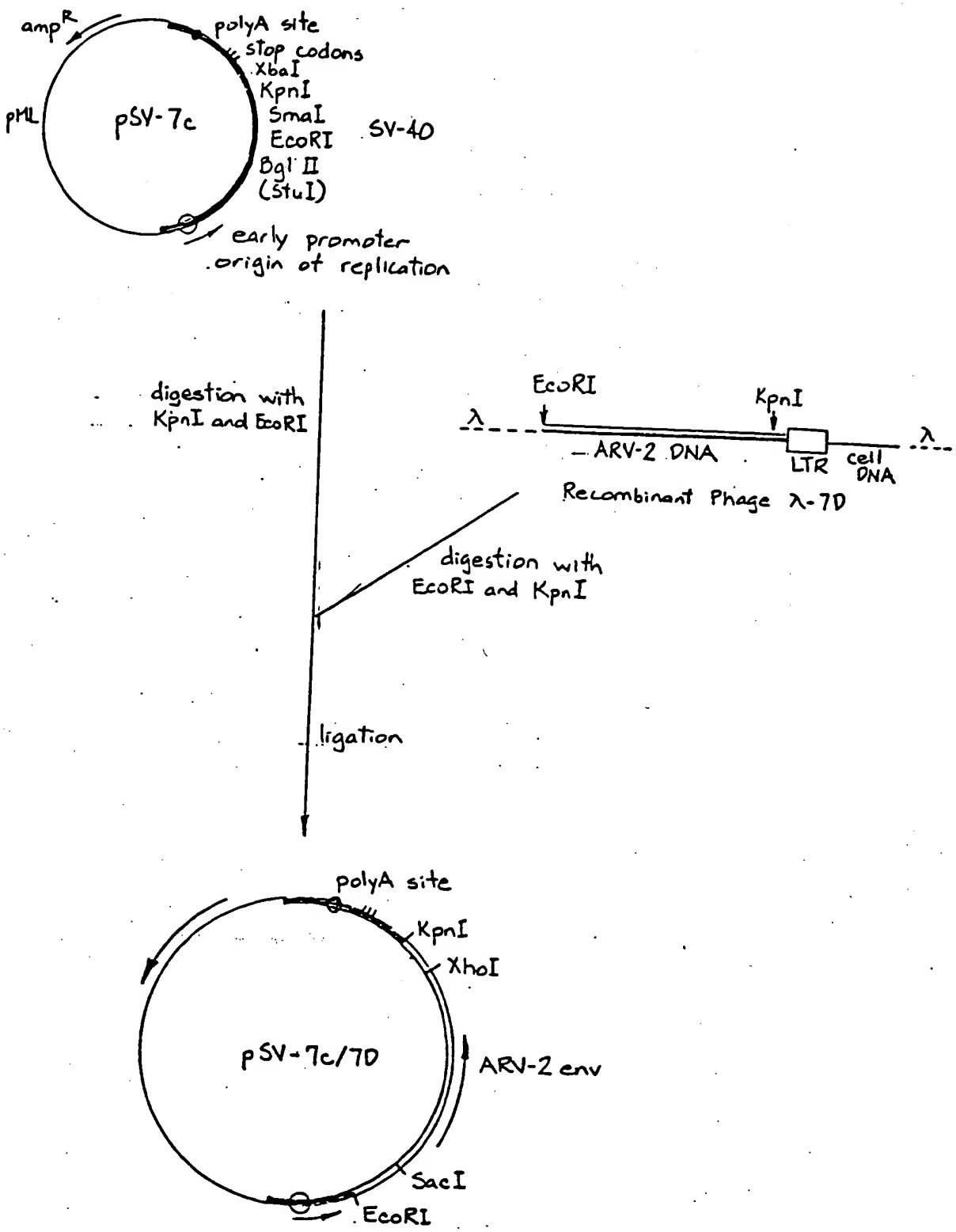

amp^R
polyA site
stop codons
XbaI
KpnI
SmaI
EcoRI
Bgl II
(StuI)
SV-40
pML
pSV-7c
early promoter
origin of replication
digestion with
KpnI and EcoRI
EcoRI
KpnI
λ
ARV-2 DNA
LTR
cell DNA
λ
Recombinant Phage λ-7D
digestion with
EcoRI and KpnI
ligation
polyA site
KpnI
XhoI
pSV-7c/7D
ARV-2 env
SacI
EcoRI

FIG. 6

Figure 7

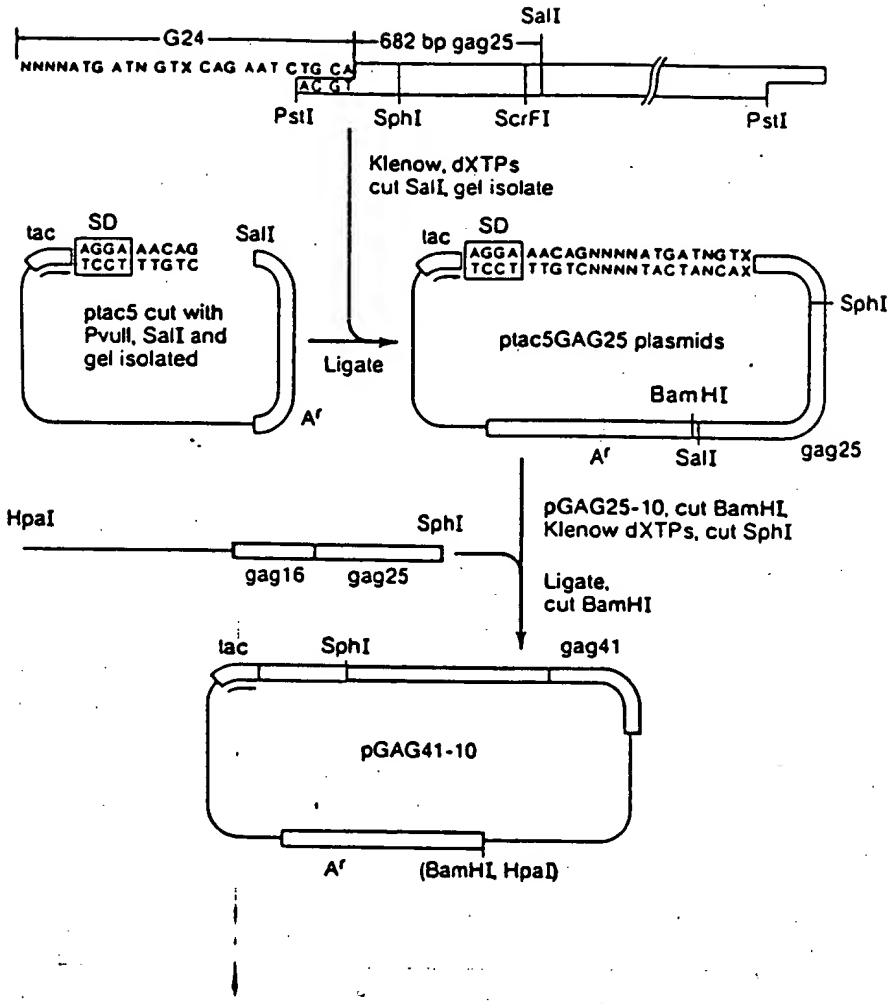
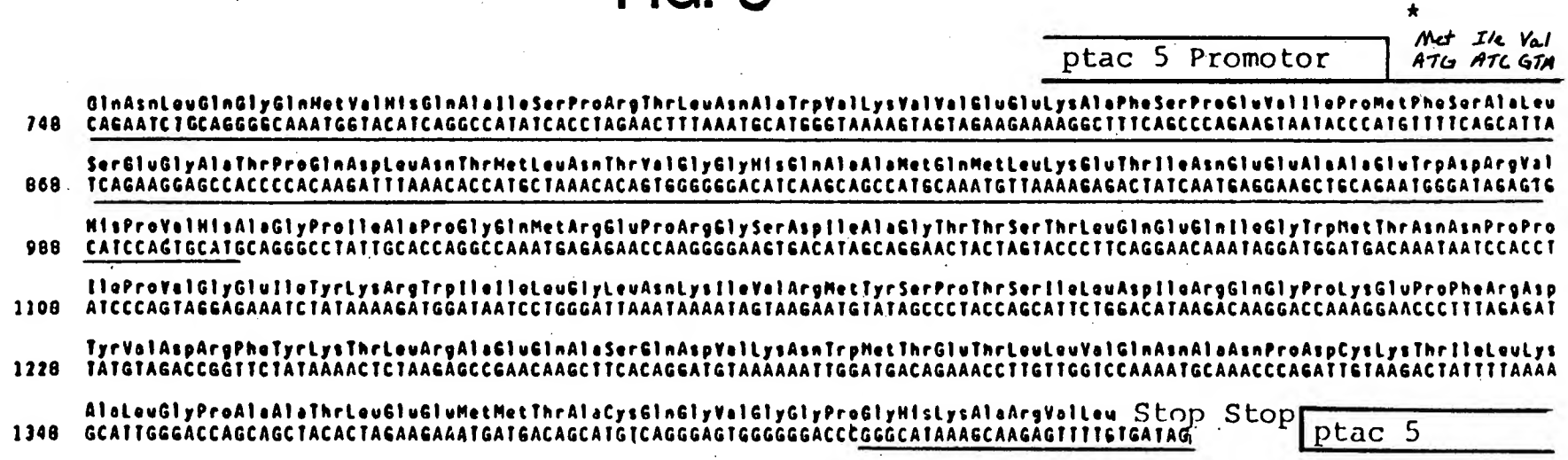

SalI
G24
682 bp gag25
NNNNATG ATN GTX CAG AAT CTG CA
AC GT
PstI
SphI
ScrFI
PstI
Klenow, dXTPs
cut SalI, gel isolate
tac
SD
AGGA
TCCT
AACAG
TTGTC
SalI
ptac5 cut with
PvuII, SalI and
gel isolated
Ar
Ligate
tac
SD
AGGA
TCCT
AACAGNNNNATGATNGTX
TTGTCNNNNTACTANCAX
SphI
ptac5GAG25 plasmids
BamHI
Ar
SalI
gag25
HpaI
SphI
gag16
gag25
pGAG25-10, cut BamHI,
Klenow dXTPs, cut SphI
Ligate,
cut BamHI
tac
SphI
gag41
pGAG41-10
Ar
(BamHI, HpaI)

FIG. 8



EP 931154

FIG. 9

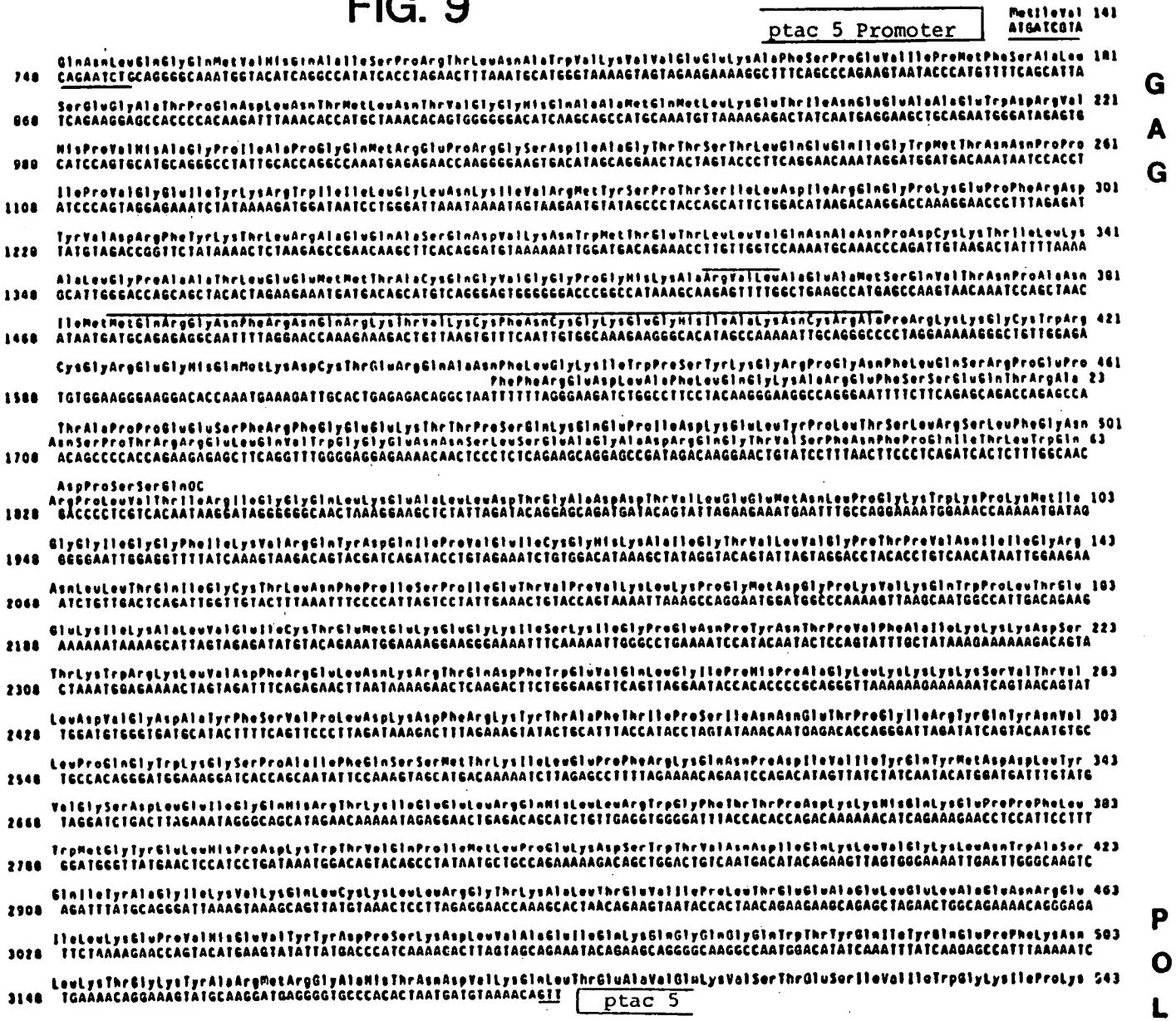
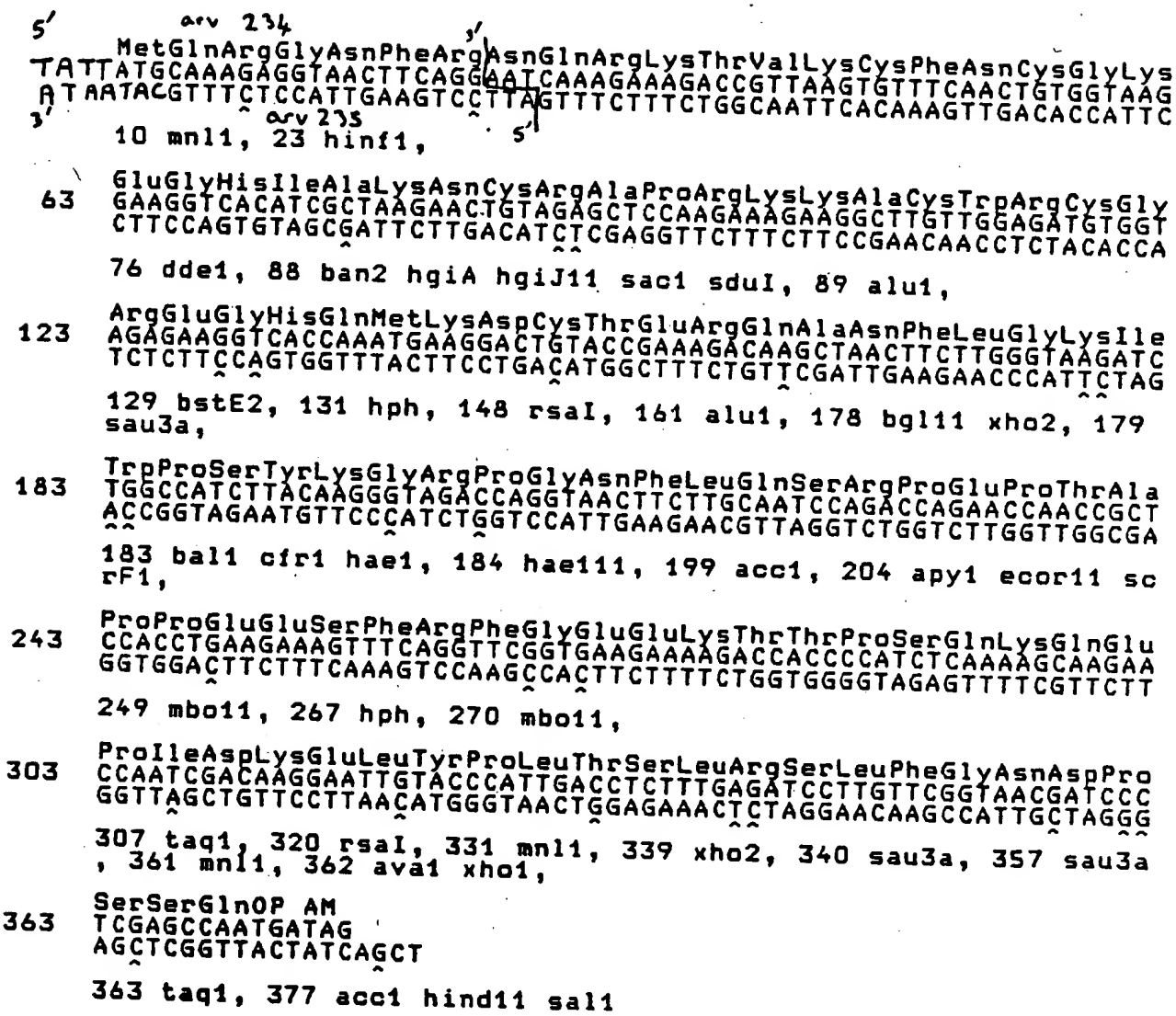


Figure 10

ARV GAG p16 - synthetic Parts A and B



PYK Promoter

```
      MetSerArgIleAspCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal 51
      ATGTCTAGAATCGATGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG

      TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGluValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal 91
5908  TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA

      ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMetGlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 131
6028  GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATGCAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA

      LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSerAsnTrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 171
6148  CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAATTGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA

      ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAspAsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 211
6268  AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGATAATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC

      IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrProAlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 251
6388  ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCGGCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA

      AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeuLeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 291
6508  AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT

      AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsnAsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 331
6628  GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAACAACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA

      IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThrLeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 371
6748  ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACTTTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT

      AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGluPhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 411
6868  AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA

      GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMetTrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 451
6988  GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA

      SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThrGluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 491
7108  TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACCGAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA

      TyrLysValIleLysIleGluProAsnSerValSer
7228  TATAAAGTAATAAAAATTGAACCAAATTCGGTATCTTGA
```

PYK Terminator

ENV

FIG. 11

NZ 931154

Nucleotide positions relative to FIGURE 5.

```
                                    MetIleAspLysAlaGlnGluGluHisGluLysTyrHisSerAsnTrp
      1 AGGXAACAG::::ATGAT:GA:AAGGCACAAGAAGAACATGAGAAATATCACAGTAATTGG
        TCCXTTGTC::::TACTA:CT:TTCCGTGTTCTTCTTGTACTCTTTATAGTGTCATTAACC

        32 mbo11, 38 nla111,

3820   62 ArgAlaMetAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSer
        AGAGCCATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC
        TCTCGGTACCGATCACTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCG

        66 nco1, 67 nla111, 118 nspBII pvu11, 119 alu1,

3880  122 CysAspLysCysGlnLeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGly
        TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA
        ACACTATTTACAGTCGATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCT

        135 alu1, 151 nla111, 152 nsi1 ava3, 155 nla111, 164 acc1, 1
        76 apy1 bstXI ecor11 scrF1,

3940  182 IleTrpGlnLeuAspCysThrHisLeuGluGlyLysIleIleLeuValAlaValHisVal
        ATATGGCAACTAGATTGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTA
        TATACCGTTGATCTAACATGTGTAGATCTTCCTTTTTAATAGGACCATCGTCAAGTACAT

        198 rsaI, 205 xba1, 223 apy1 ecor11 scrF1, 236 nla111,

4000  242 AlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyr
        GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATAT
        CGGTCACCTATATATCTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATA

        263 xmn1,

4060  302 PheLeuLeuLysLeuAlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySer
        TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGC
        AAAGAGAATTTTAATCGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCG

        321 mbo11, 326 bal1 cfr1 hae1, 327 hae111, 357 bbv fnu4h1,

4120  362 AsnPheThrSerThrThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPhe
        AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTT
        TTAAAGTGGTCATGATGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAA

        366 hph, 371 sca1, 372 rsaI, 385 hae111, 386 fnu4h1 nsb11, 4
        05 binI, 406 dpn1 sau3a,

4180  422 GlyIleProTyrAsnProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLys
        GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAG
        CCGTAAGGGATGTTAGGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTC

        423 bsm1, 458 hinf1,

4240  482 LysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAla
        AAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCA
        TTTTAATATCCTGTCCATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGT

        503 dpn1 sau3a, 518 afl11, 530 rsaI,

4300  542 ValPheIleHisAsnPheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArg
        GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGA
        CATAAGTAGGTGTTAAAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCT

        547 fok1, 557 aha111,

4360  602 IleValAspIleIleAlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLys
        ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAA
        TATCATCTGTATTATCGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTT

        605 acc1,

      662 IleGlnAsnPheArgValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAla
        ATTCAAAATTTTCGGGTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCA
```

```
                LysLeuLeuTrpLysGlyGluGlyAlaValIleGlnAspAsnSerAspIleLysVal
4480    722     AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA
                TTCGAAGAGACCTTTCCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCAT
                  ^^  ^                                                  ^

                722 hind111, 723 alu1, 737 hph,

                ValProArgArgLysAlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAsp
4540    782     GTGCCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGAT
                CACGGTTCTTCTTTTCGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTA
                       ^                                        ^

                789 mbo11, 833 hph,

                CysValAlaSerArgGlnAspGluAspAM
4600    842     TGTGTGGCAAGTAGACAGGATGAGGATTAGTCGACGGAATTCTTTAGTAAAACACC
                ACACACCGTTCATCTGTCCTACTCCTAATCAGCTGCCTTAAGAAATCATTTTGTGG
                       ^        ^  ^  ^^      ^

                852 acc1, 859 fok1, 863 mnl1, 871 acc1 hind11 sal1, 872 taq1
                , 878 ecor1,
```

FIGURE 12

2 of 2

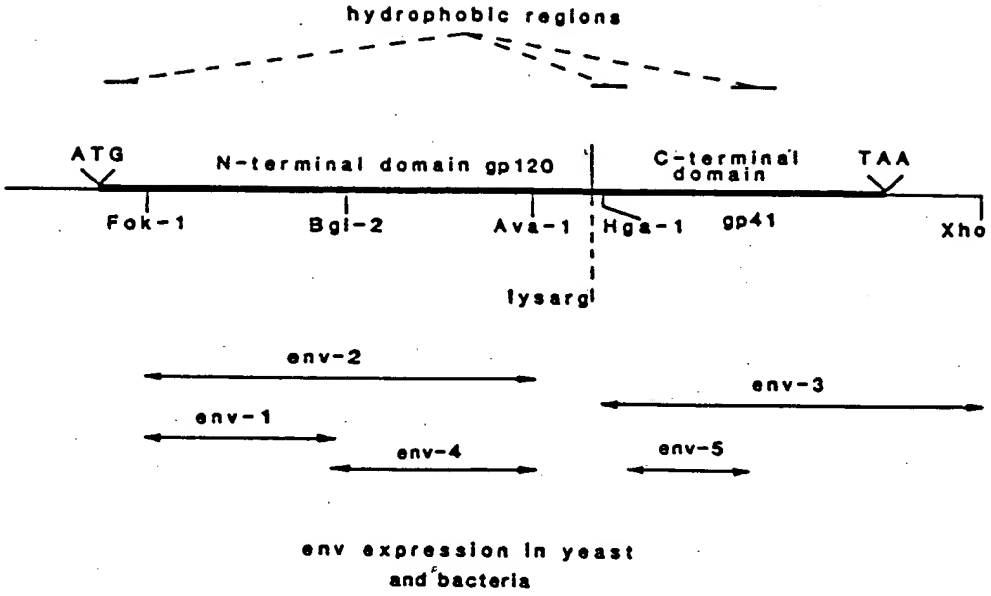

hydrophobic regions
ATG
N-terminal domain gp120
C-terminal domain
TAA
Fok-1
Bgl-2
Ava-1
Hga-1
gp41
Xho
lysarg
env-2
env-3
env-1
env-4
env-5
env expression in yeast and bacteria

FIGURE 13

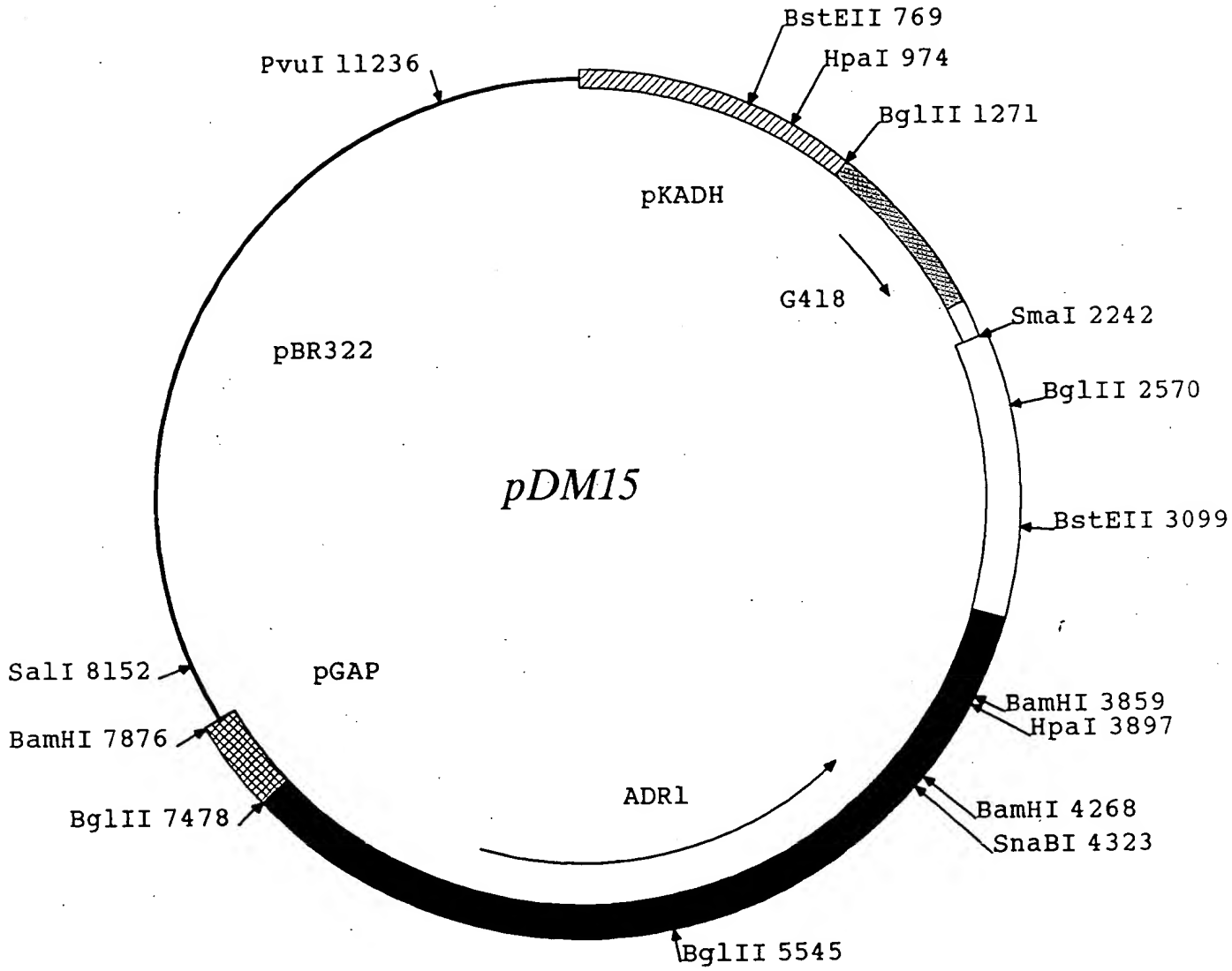

BstEII 769
HpaI 974
PvuI 11236
BglII 1271
pKADH
G418
SmaI 2242
pBR322
BglII 2570
pDM15
BstEII 3099
SalI 8152
pGAP
BamHI 3859
HpaI 3897
BamHI 7876
ADR1
BglII 7478
BamHI 4268
SnaBI 4323
BglII 5545

FIGURE 14

SOD

```
    MetAlaThrLysAlaValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsn
  1 CATGGCGACGAAGGCCGTGTGCGTGCTGAAGGGCGACGGCCCAGTGCAGGGCATCATCAAT
    GGCTGCTTCCGGCACACGCACGACTTCCCGCTGCCGGGTCACGTCCCGTAGTAGTTA

    PheGluGlnLysGluSerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThr
 62 TTCGAGCAGAAGGAAAGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACT
    AAGCTCGTCTTCCTTTCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGA

    GluGlyLeuHisGlyPheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSer
122 GAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGT
    CTTCCGGACGTACCTAAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCA

    AlaGlyProHisPheAsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArg
182 GCAGGTCCTCACTTTAATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGG
    CGTCCAGGAGTGAAATTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCC

    HisValGlyAspLeuGlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIle
242 CATGTTGGAGACTTGGGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATT
    GTACAACCTCTGAACCCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAA

    GluAspSerValIleSerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValVal
302 GAAGATTCTGTGATCTCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTC
    CTTCTAAGACACTAGAGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAG

    HisGluLysAlaAspAspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsn
362 CATGAAAAAGCAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAAC
    GTACTTTTTCGTCTACTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTG

                                                    ENV 5B
    AlaGlySerArgLeuAlaCysGlyValIleGlyIleAlaMetAlaIleGluAlaGlnGln
422 GCTGGAAGTCGTTTGGCTTGTGGTGTAATTGGGATCGCCATGGCTATCGAAGCTCAACAA
    CGACCTTCAGCAAACCGAACACCACATTAACCCTAGCGGTACCGATAGCTTCGAGTTGTT
                                                    401

    HisLeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaVal
482 CACTTGCTGCAGTTGACCGTTTGGGGTATCAAGCAGTTGCAGGCTAGAGTTTTGGCTGTT
    GTGAACGACGTCAACTGGCAAACCCCATAGTTCGTCAACGTCCGATCTCAAAACCGACAA

    GluArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIle
542 GAAAGATACTTGAGAGATCAACAATTGTTGGGTATCTGGGGTTGTTCTGGTAAGTTGATT
    CTTTCTATGAACTCTCTAGTTGTTAACAACCCATAGACCCCAACAAGACCATTCAACTAA

    CysThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrp
602 TGTACCACCGCTGTTCCCTGGAACGCTTCTTGGTCTAACAAGTCTTTGGAAGACATCTGG
    ACATGGTGGCGACAAGGGACCTTGCGAAGAACCAGATTGTTCAGAAACCTTCTGTAGACC

    AspAsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyr
662 GACAACATGACCTGGATGCAATGGGAAAGAGAAATCGACAACTACACCAACACCATCTAC
    CTGTTGTACTGGACCTACGTTACCCTTTCTCTTTAGCTGTTGATGTGGTTGTGGTAGATG

    ThrLeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeu
722 ACCTTGTTGGAGGAATCTCAAAACCAACAAGAAAAGAACGAACAAGAATTGTTGGAATTG
    TGGAACAACCTCCTTAGAGTTTTGGTTGTTCTTTTCTTGCTTGTTCTTAACAACCTTAAC

    AspLysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpAM
782 GACAAGTGGGCAAGCTTGTGGAACTGGTTCTCTATCACCAACTGGTAG
    CTGTTCACCCGTTCGAACACCTTGACCAAGAGATAGTGGTTGACCATCAGCT
```

Translated Mol. Weight = 30414.22

FIGURE 15

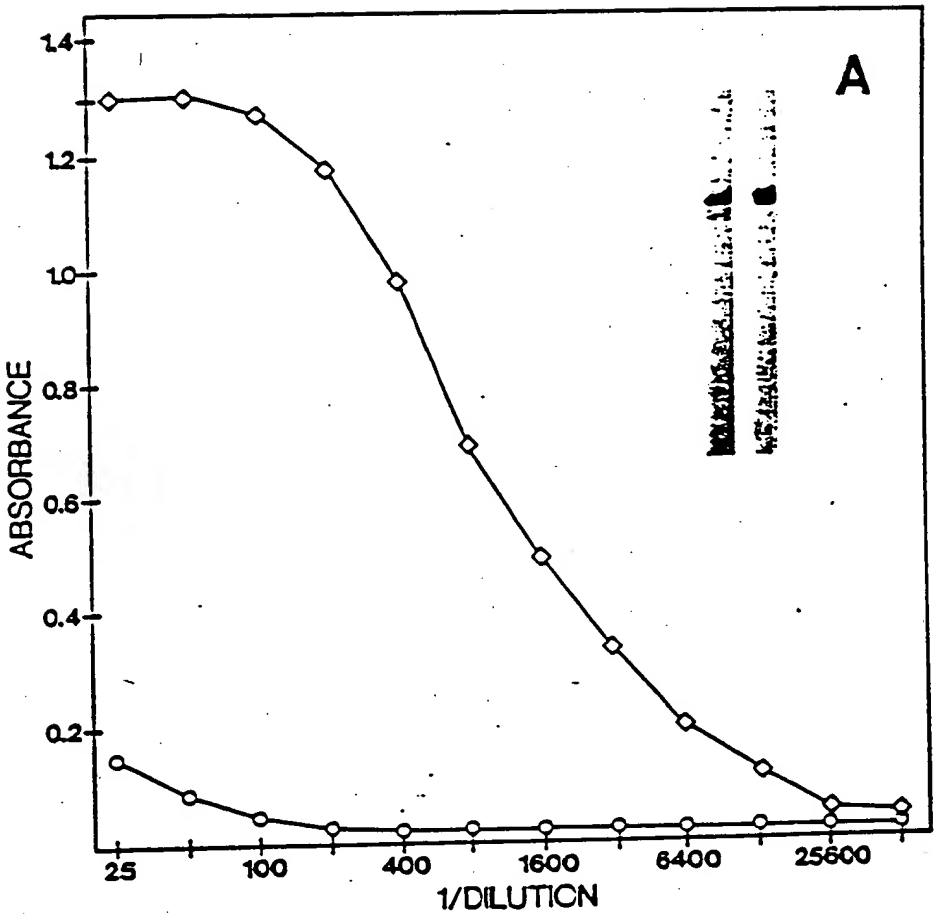

A
ABSORBANCE
1.4
1.2
1.0
0.8
0.6
0.4
0.2
25
100
400
1600
6400
25600
1/DILUTION

FIGURE 16

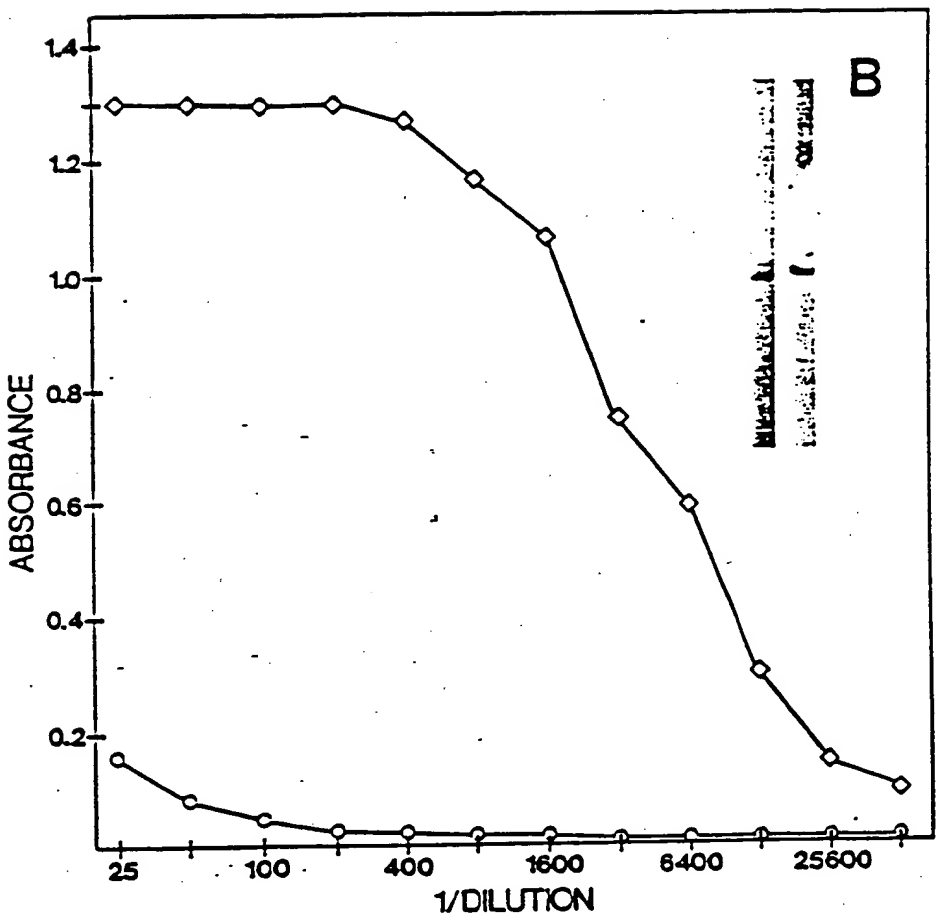

B
ABSORBANCE
1.4
1.2
1.0
0.8
0.6
0.4
0.2
25
100
400
1600
6400
25600
1/DILUTION

FIGURE 16

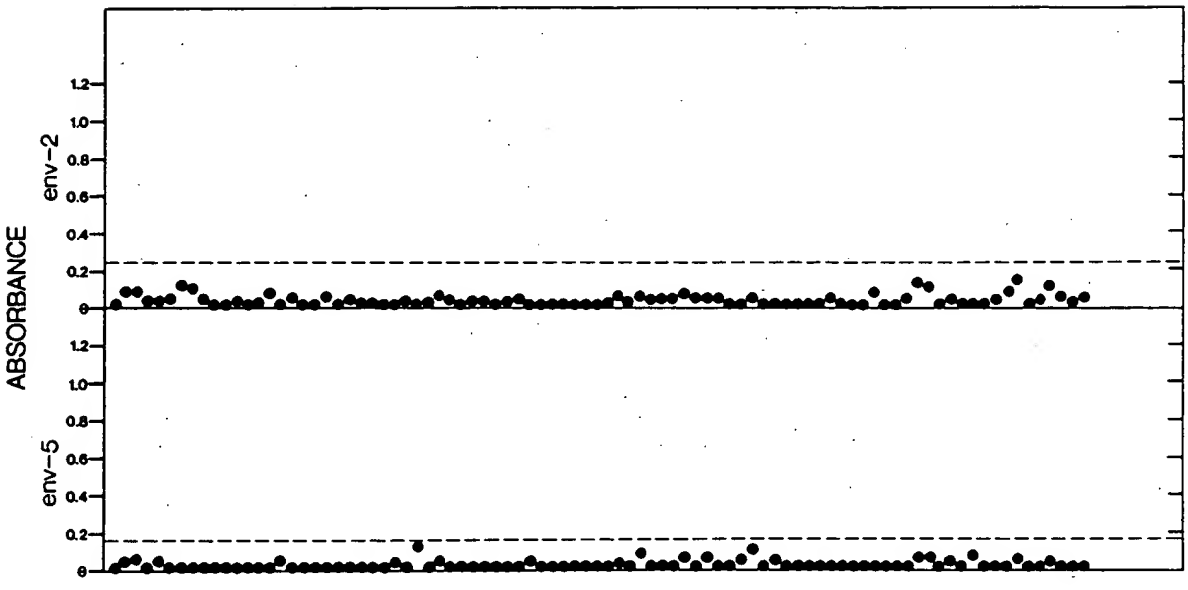


FIGURE 17

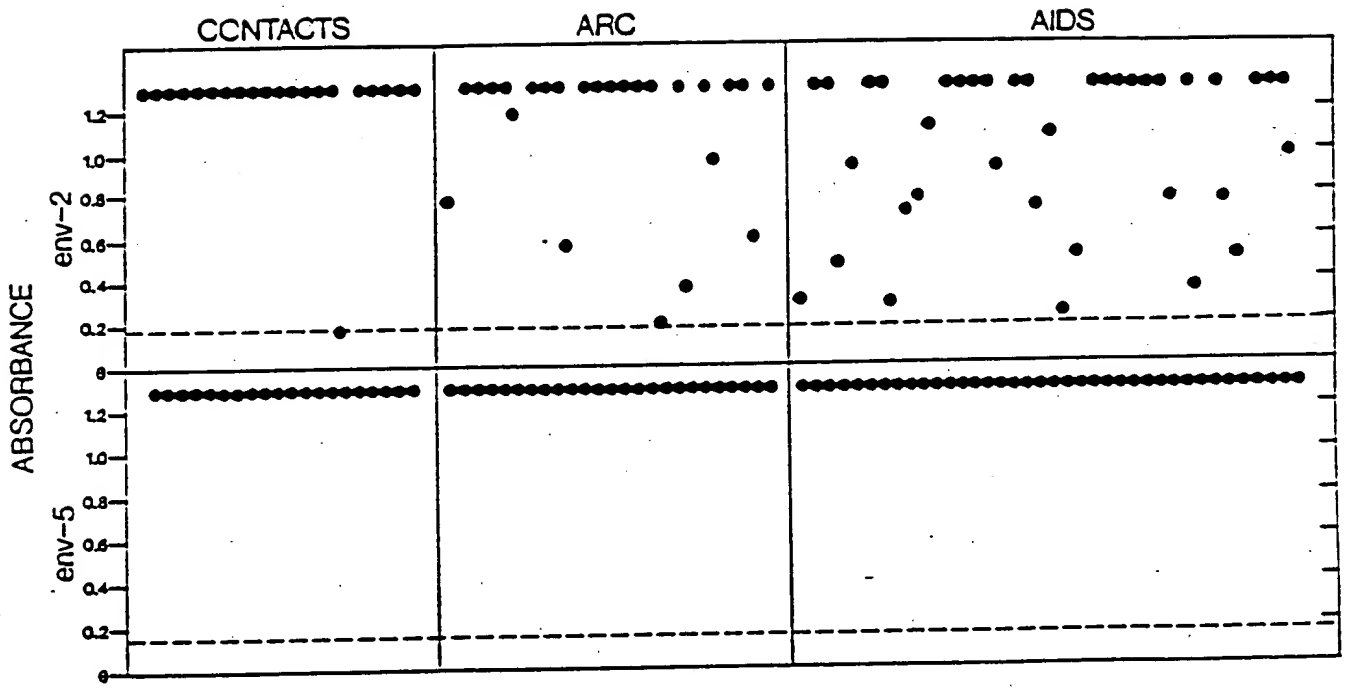


FIGURE 18

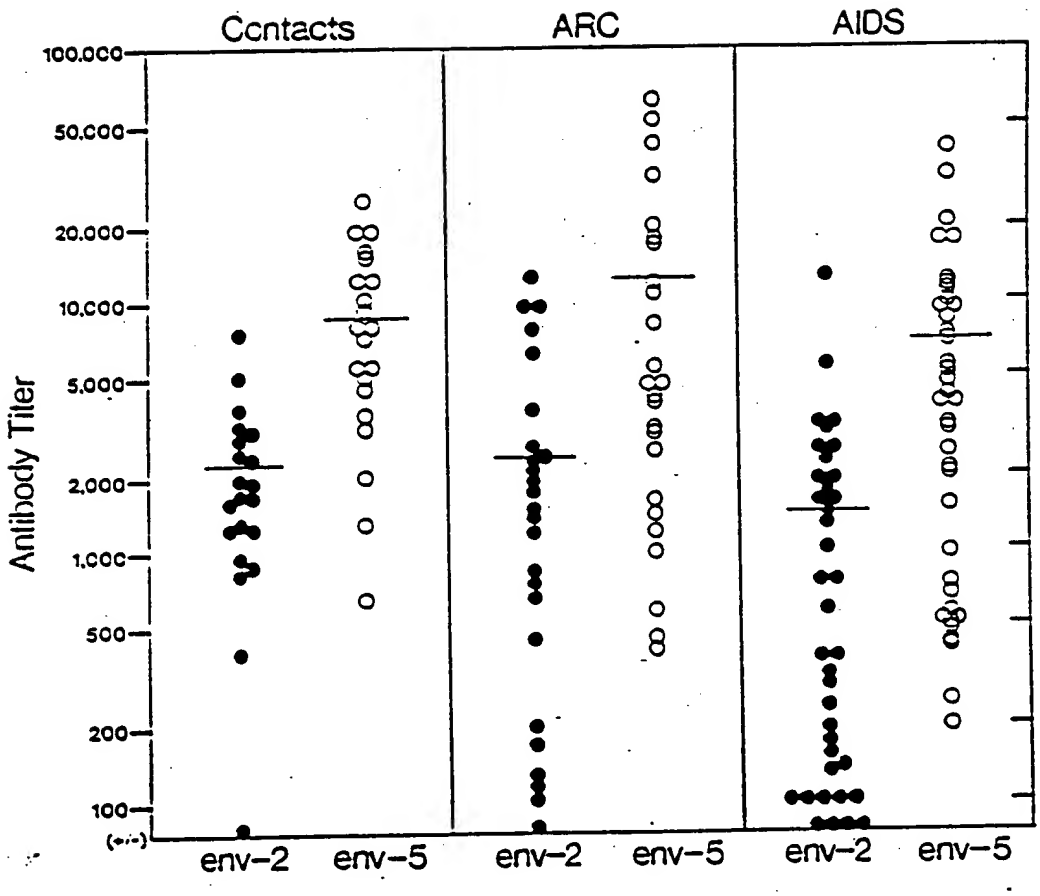

Contacts
ARC
AIDS
Antibody Titer
100,000
50,000
20,000
10,000
5,000
2,000
1,000
500
200
100
(+/-)
env-2
env-5
env-2
env-5
env-2
env-5

FIGURE 19

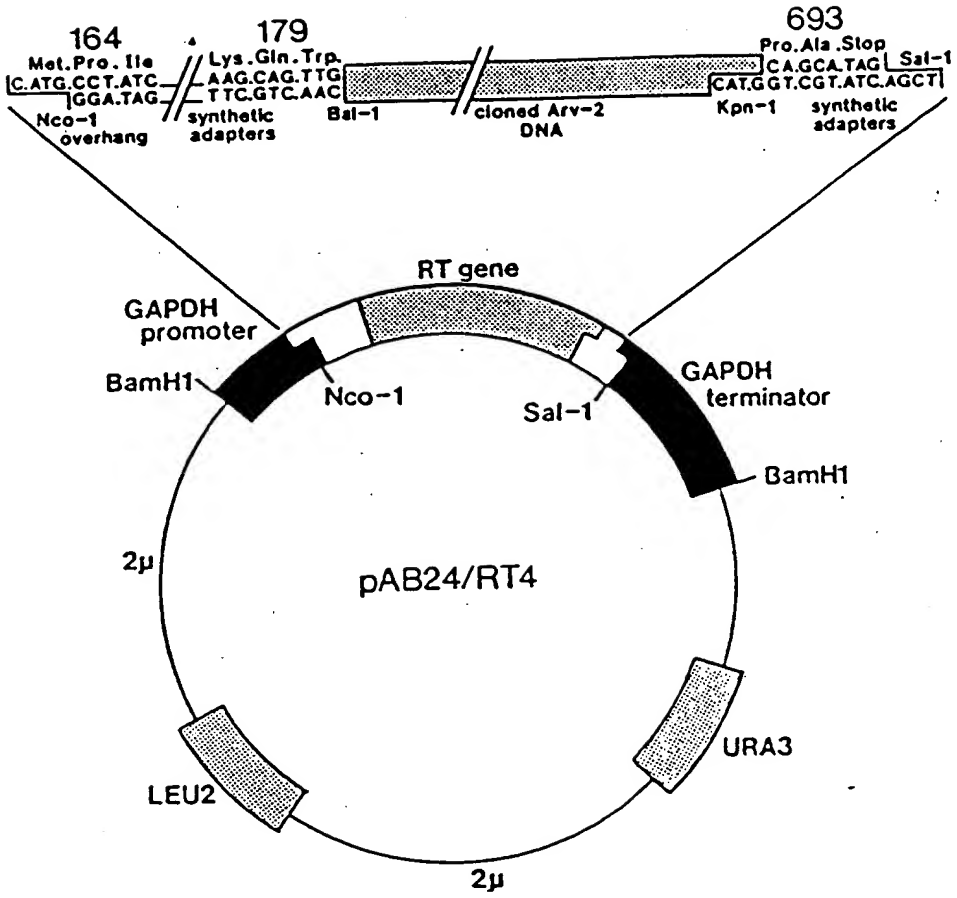

164
Met.Pro.Ile
C.ATG.CCT.ATC
GGA.TAG
Nco-1
overhang
179
Lys.Gln.Trp.
AAG.CAG.TTG
TTC.GTC.AAC
synthetic
adapters
Bal-1
cloned Arv-2
DNA
693
Pro.Ala.Stop
CA.GCA.TAG
Sal-1
CAT.GGT.CGT.ATC.AGCT
Kpn-1
synthetic
adapters
RT gene
GAPDH
promoter
BamH1
Nco-1
Sal-1
GAPDH
terminator
BamH1
2μ
pAB24/RT4
URA3
LEU2
2μ

FIGURE 20

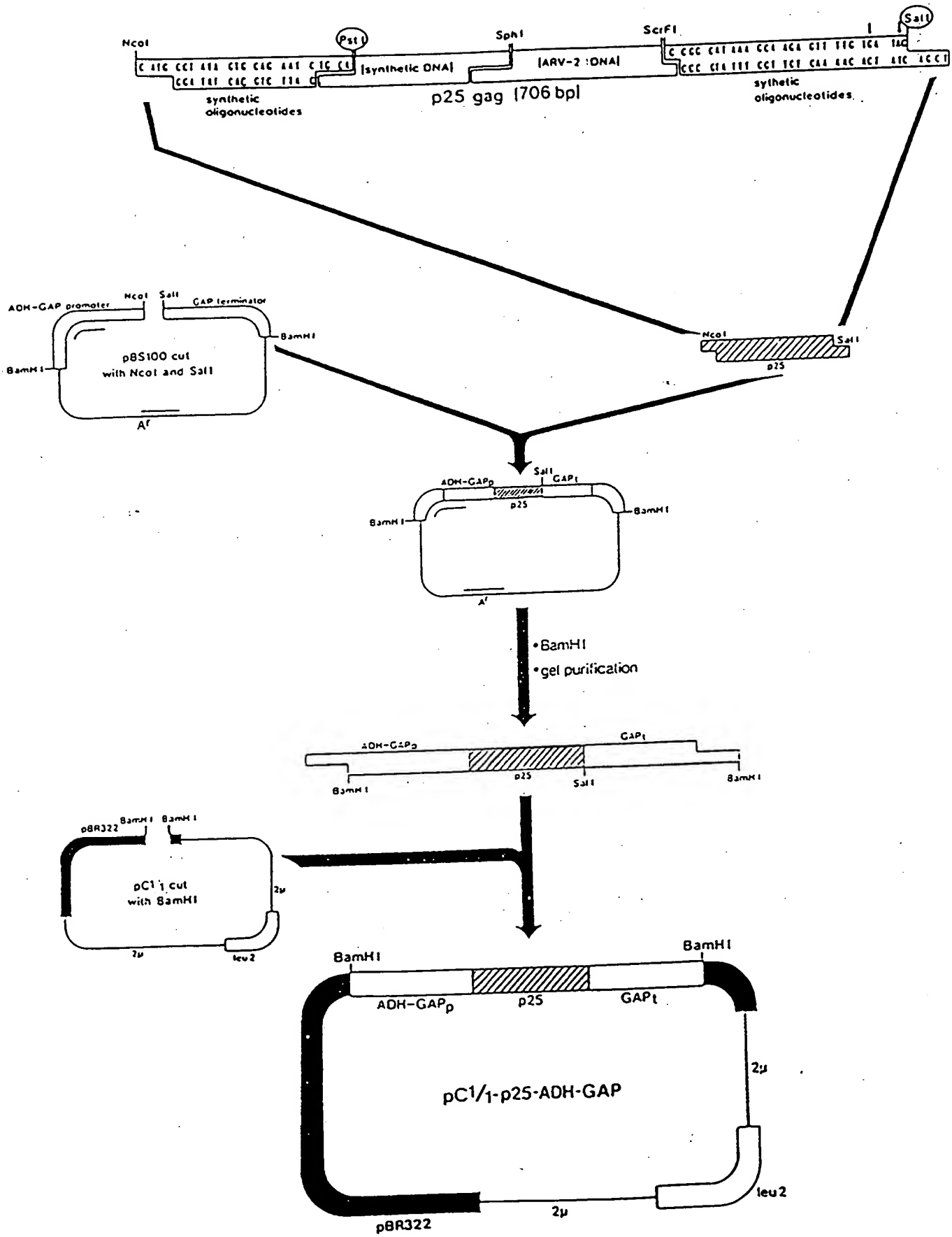

NcoI
PstI
SphI
ScrFI
SalI
synthetic DNA
ARV-2 DNA
synthetic oligonucleotides
p25 gag 706 bpl
synthetic oligonucleotides
ADH-GAP promoter
NcoI
SalI
GAP terminator
BamHI
BamHI
pBS100 cut with NcoI and SalI
Ap
NcoI
SalI
p25
ADH-GAPp
SalI
GAPt
p25
BamHI
BamHI
Ap
•BamHI
•gel purification
ADH-GAPp
GAPt
p25
SalI
BamHI
BamHI
pBR322
BamHI
BamHI
pC1/1 cut with BamHI
2μ
2μ
leu2
BamHI
BamHI
ADH-GAPp
p25
GAPt
pC1/1-p25-ADH-GAP
2μ
leu2
2μ
pBR322

FIGURE 21

```
     1                                                  10
     Met Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln
C    ATG CCT ATA GTG CAG AAT CTG CAG GGG CAA ATG GTA CAT CAG

                        20
Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
GCC ATA TCA CCT AGA ACT TTA AAT GCT TGG GTA AAA GTA GTA GAA

 30                                                 40
Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu
GAA AAG GCT TTC AGC CCA GAA GTA ATA CCC ATG TTT TCA GCA TTA

                        50
Ser Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr
TCA GAA GGA GCC ACC CCT CAA GAT TTA AAC ACC ATG CTA AAC ACA

 60                                                 70
Val Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile
GTG GGG GGA CAT CAA GCA GCC ATG CAA ATG TTA AAA GAG ACT ATC

                        80
Asn Glu Glu Ala Ala Glu Trp Asp Arg Val His Pro Val His Ala
AAT GAG GAG GCT GCC GAA TGG GAT AGA GTG CAT CCA GTG CAT GCA

 90                                                100
Gly Pro Ile Ala Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp
GGG CCT ATT GCA CCA GGC CAA ATG AGA GAA CCA AGG GGA AGT GAC

                       110
Ile Ala Gly Thr Thr Ser Thr Leu Gln Glu Gln Ile Gly Trp Met
ATA GCA GGA ACT ACT AGT ACC CTT CAG GAA CAA ATA GGA TGG ATG

120                                                130
Thr Asn Asn Pro Pro Ile Pro Val Gly Glu Ile Tyr Lys Arg Trp
ACA AAT AAT CCA CCT ATC CCA GTA GGA GAA ATC TAT AAA AGA TGG

                       140
Ile Ile Leu Gly Leu Asn Lys Ile Val Arg Met Tyr Ser Pro Thr
ATA ATC CTG GGA TTA AAT AAA ATA GTA AGA ATG TAT AGC CCT ACC

150                                                160
Ser Ile Leu Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp
AGC ATT CTG GAC ATA AGA CAA GGA CCA AAG GAA CCC TTT AGA GAT

                       170
Tyr Val Asp Arg Phe Tyr Lys Thr Leu Arg Ala Glu Gln Ala Ser
TAT GTA GAC CGG TTC TAT AAA ACT CTA AGA GCC GAA CAA GCT TCA

180                                                190
Gln Asp Val Lys Asn Trp Met Thr Glu Thr Leu Leu Val Gln Asn
CAG GAT GTA AAA AAT TGG ATG ACA GAA ACC TTG TTG GTC CAA AAT

                       200
Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala Leu Gly Pro Ala
GCA AAC CCA GAT TGT AAG ACT ATT TTA AAA GCA TTG GGA CCA GCA

210                                                220
Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
GCT ACA CTA GAA GAA ATG ATG ACA GCA TGT CAG GGA GTG GGG GGA

                       230     232
Pro Gly His Lys Ala Arg Val Leu OP
CCC GGG CAT AAA GCA AGA GTT TTG TGA TAG

Translated Mol. Weight = 25700.75
```

FIGURE 22

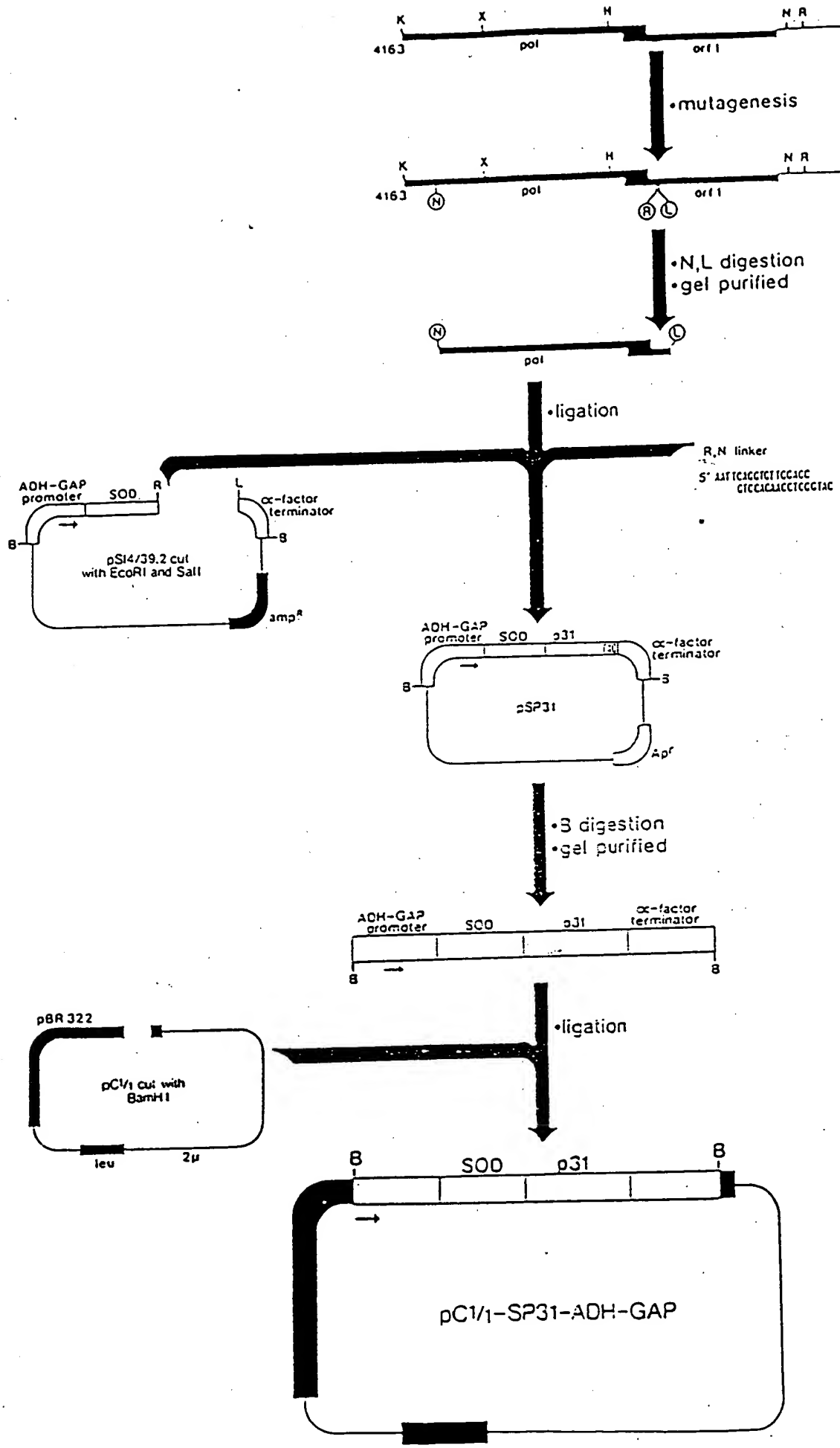

K
X
H
N R
T
ARV-2
4163
6011
pol
orf 1
•mutagenesis
•N,L digestion
•gel purified
•ligation
R,N linker
ADH-GAP promoter
SOD
α-factor terminator
pSI4/39.2 cut with EcoRI and SalI
amp^R
p31
pSP31
Ap^r
•B digestion
•gel purified
pBR 322
pC1/1 cut with BamHI
leu
2μ
pC1/1-SP31-ADH-GAP

FIGURE 23

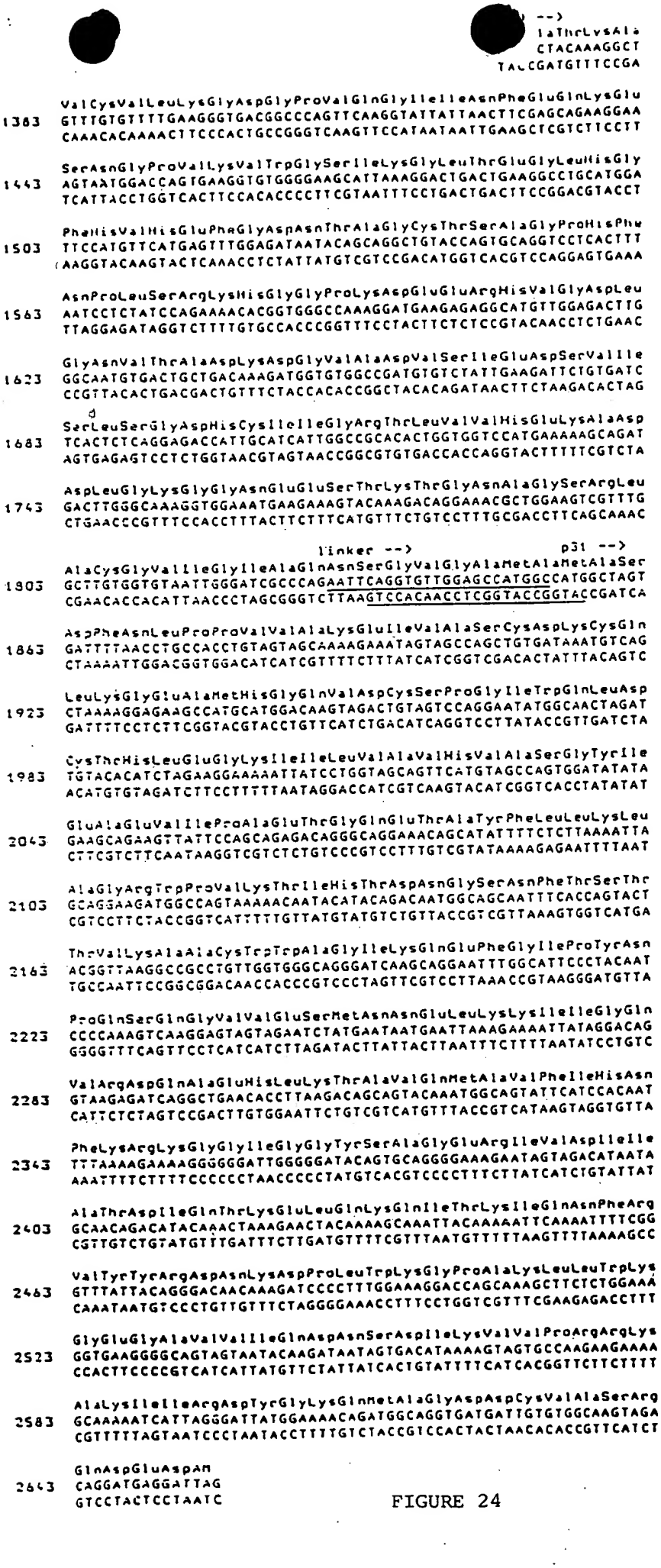

--> laThrLysAla CTACAAAGGCT TALCGATGTTTCCGA
ValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsnPheGluGlnLysGlu
1383 GTTTGTGTTTTGAAGGGTGACGGCCCAGTTCAAGGTATTATTAACTTCGAGCAGAAGGAA
CAAACACAAAACTTCCCACTGCCGGGTCAAGTTCCATAATAATTGAAGCTCGTCTTCCTT
SerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThrGluGlyLeuHisGly
1443 AGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACTGAAGGCCTGCATGGA
TCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGACTTCCGGACGTACCT
PheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSerAlaGlyProHisPhe
1503 TTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGTGCAGGTCCTCACTTT
AAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCACGTCCAGGAGTGAAA
AsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArgHisValGlyAspLeu
1563 AATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGGCATGTTGGAGACTTG
TTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCCGTACAACCTCTGAAC
GlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIleGluAspSerValIle
1623 GGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATTGAAGATTCTGTGATC
CCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAACTTCTAAGACACTAG
SerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValValHisGluLysAlaAsp
1683 TCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTCCATGAAAAAGCAGAT
AGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAGGTACTTTTTCGTCTA
AspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsnAlaGlySerArgLeu
1743 GACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAACGCTGGAAGTCGTTTG
CTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTGCGACCTTCAGCAAAC
linker --> p31 -->
AlaCysGlyValIleGlyIleAlaGlnAsnSerGlyValGlyAlaMetAlaMetAlaSer
1803 GCTTGTGGTGTAATTGGGATCGCCCAGAATTCAGGTGTTGGAGCCATGGCCATGGCTAGT
CGAACACCACATTAACCCTAGCGGGTCTTAAGTCCACAACCTCGGTACCGGTACCGATCA
AspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLysCysGln
1863 GATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAG
CTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTC
LeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAsp
1923 CTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGAT
GATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTA
CysThrHisLeuGluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIle
1983 TGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATA
ACATGTGTAGATCTTCCTTTTTAATAGGACCATCGTCAAGTACATCGGTCACCTATATAT
GluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeu
2043 GAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTA
CTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATAAAAGAGAATTTTAAT
AlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThr
2103 GCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACT
CGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATGA
ThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsn
2163 ACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAAT
TGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAACCGTAAGGGATGTTA
ProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGln
2223 CCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAG
GGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTCTTTTAATATCCTGTC
ValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsn
2283 GTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAAT
CATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTA
PheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIle
2343 TTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATA
AAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCTTATCATCTGTATTAT
AlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArg
2403 GCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGG
CGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTTTAAGTTTTAAAAGCC
ValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLys
2463 GTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAA
CAAATAATGTCCCTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGTTTCGAAGAGACCTTT
GlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysValValProArgArgLys
2523 GGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAA
CCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTT
AlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArg
2583 GCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGA
CGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCT
GlnAspGluAspAm
2643 CAGGATGAGGATTAG
GTCCTACTCCTAATC

FIGURE 24

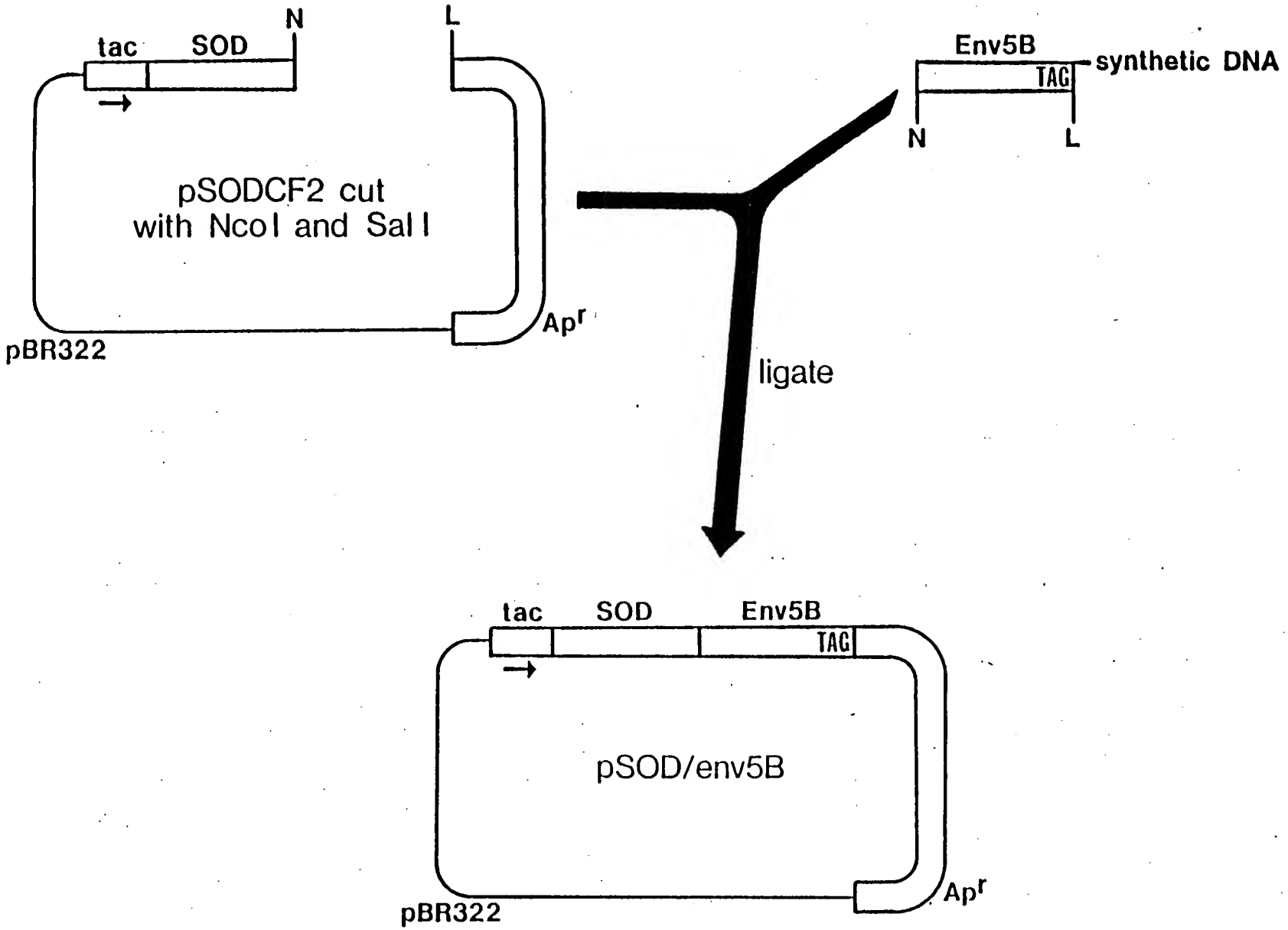
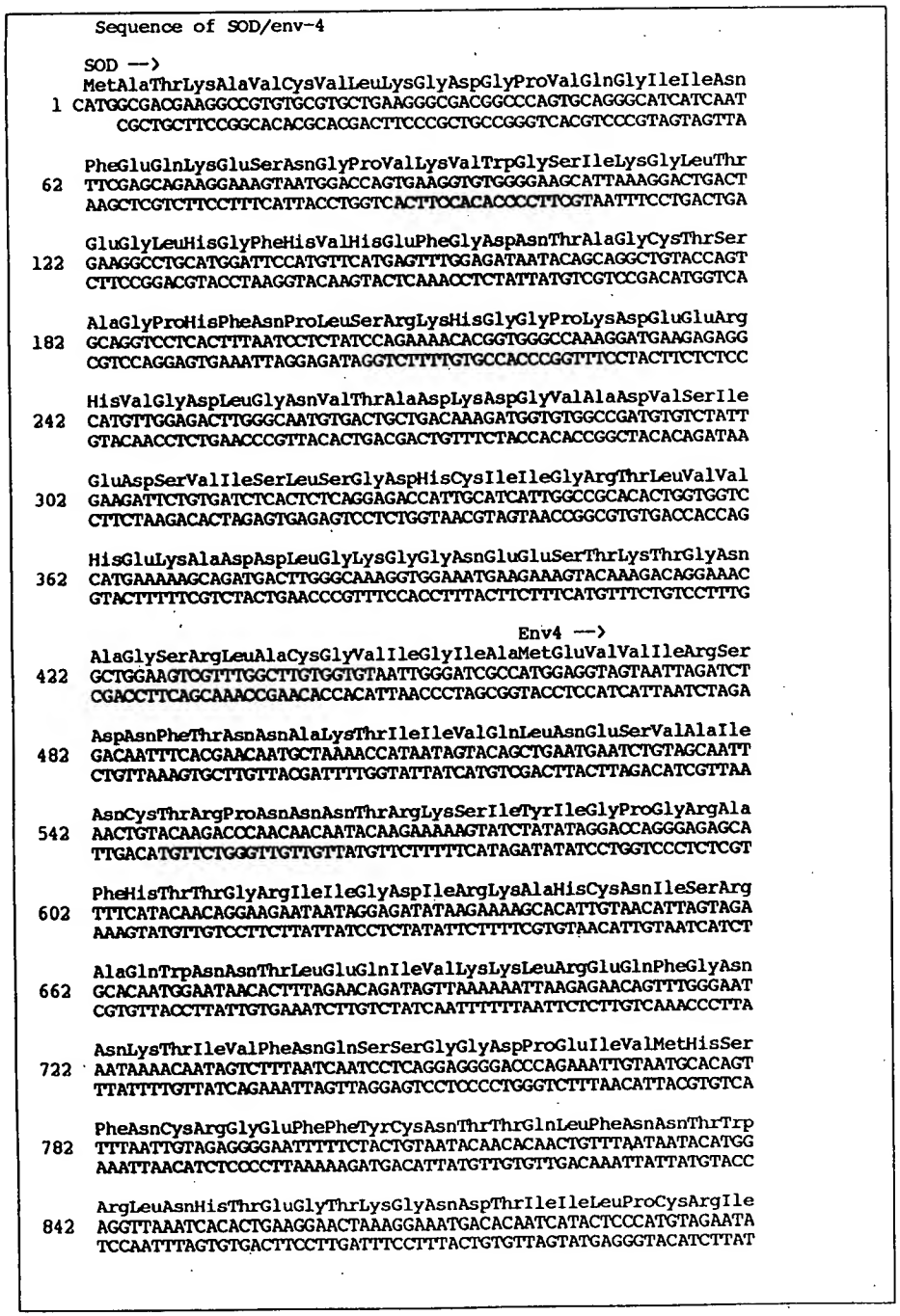
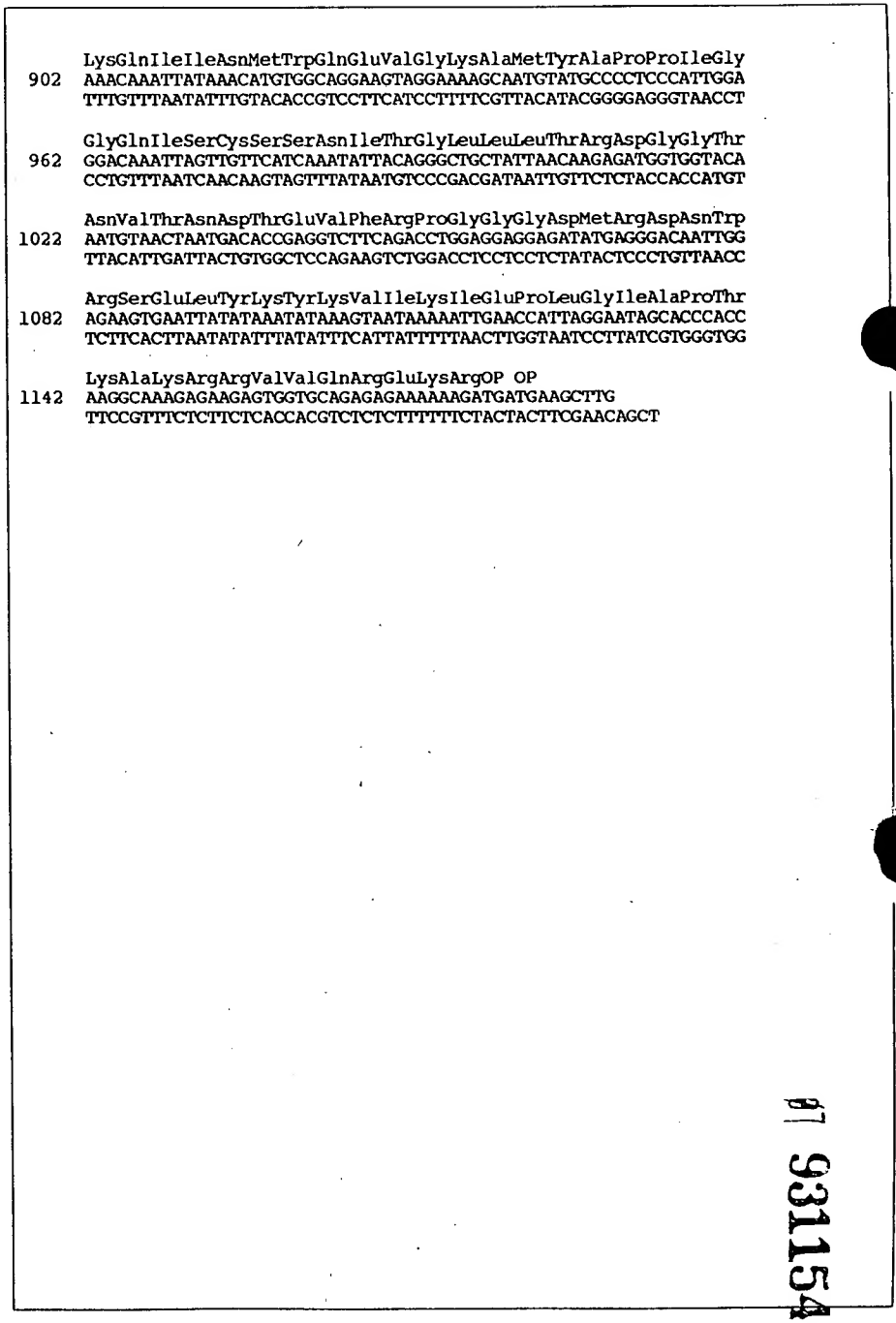


FIGURE 25

FIGURE 26





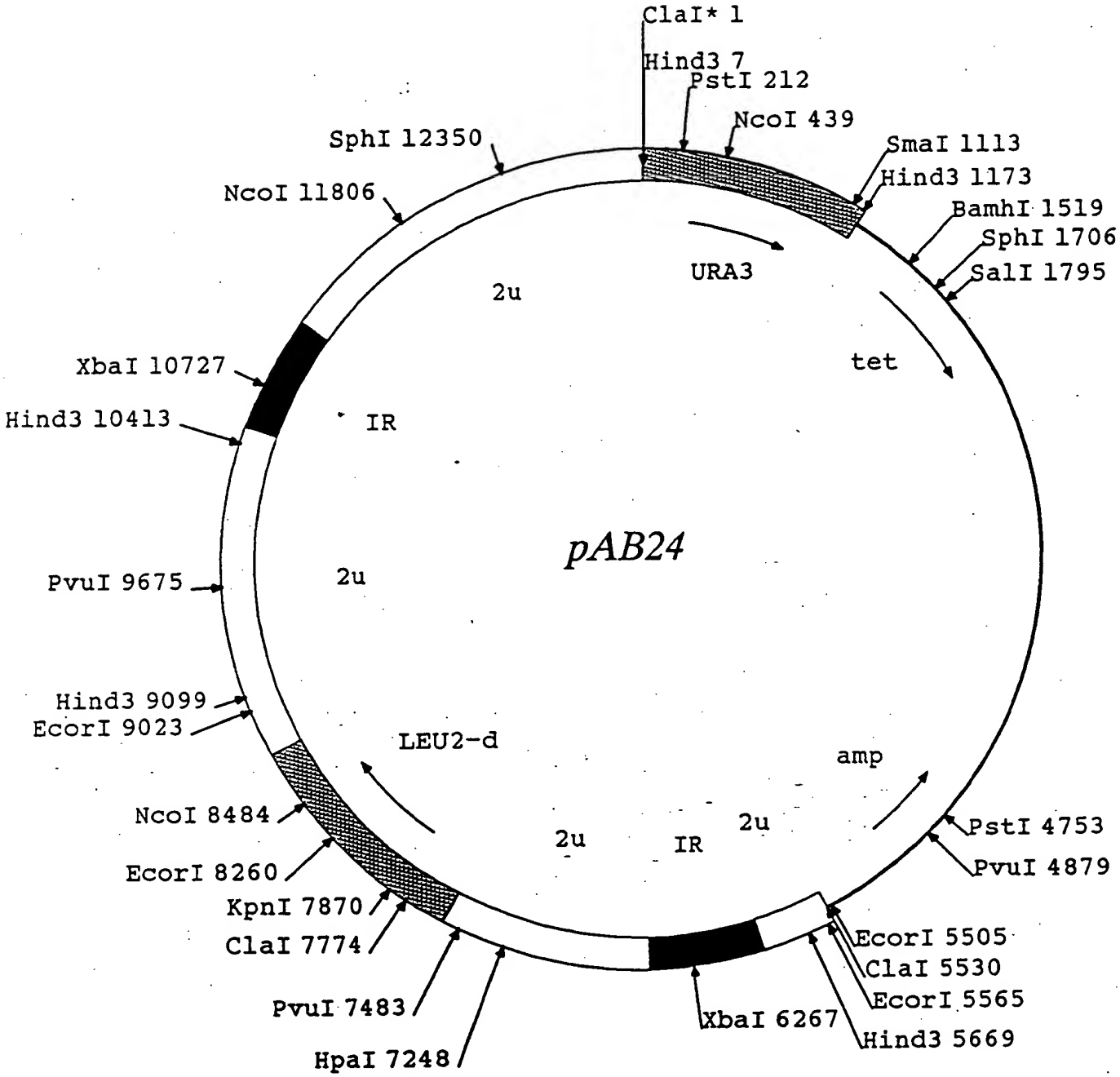

ClaI* 1
Hind3 7
PstI 212
NcoI 439
SmaI 1113
Hind3 1173
BamhI 1519
SphI 1706
SalI 1795
SphI 12350
NcoI 11806
XbaI 10727
Hind3 10413
PvuI 9675
Hind3 9099
EcorI 9023
NcoI 8484
EcorI 8260
KpnI 7870
ClaI 7774
PvuI 7483
HpaI 7248
XbaI 6267
EcorI 5505
ClaI 5530
EcorI 5565
Hind3 5669
PstI 4753
PvuI 4879
URA3
tet
amp
LEU2-d
2u
2u
2u
IR
IR
pAB24

FIGURE 27

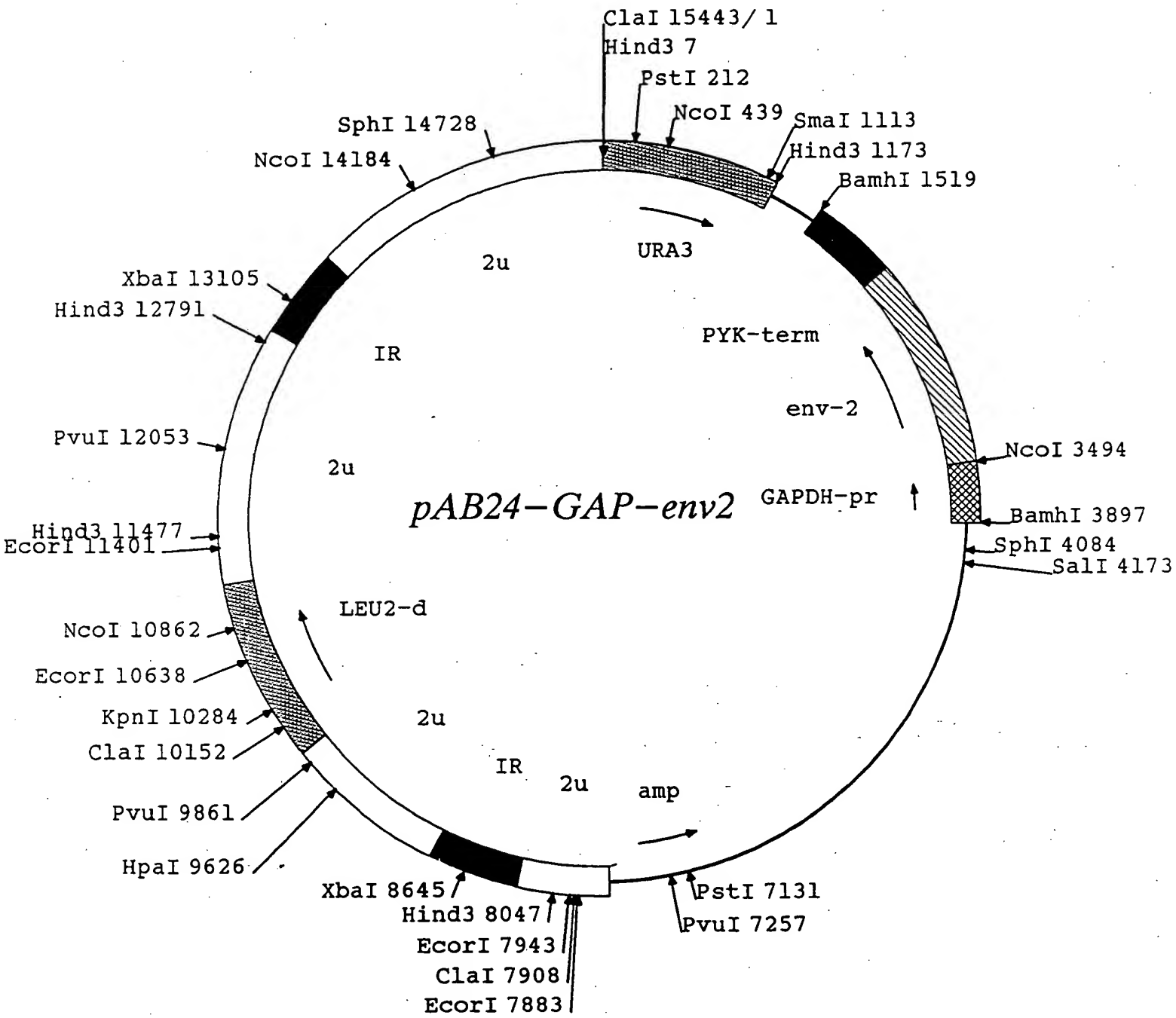

ClaI 15443/ 1
Hind3 7
PstI 212
NcoI 439
SmaI 1113
Hind3 1173
BamhI 1519
SphI 14728
NcoI 14184
URA3
2u
XbaI 13105
Hind3 12791
IR
PYK-term
env-2
PvuI 12053
2u
NcoI 3494
pAB24-GAP-env2
GAPDH-pr
BamhI 3897
Hind3 11477
EcorI 11401
SphI 4084
SalI 4173
LEU2-d
NcoI 10862
EcorI 10638
KpnI 10284
2u
ClaI 10152
IR
2u
amp
PvuI 9861
HpaI 9626
XbaI 8645
PstI 7131
Hind3 8047
PvuI 7257
EcorI 7943
ClaI 7908
EcorI 7883

FIGURE 28

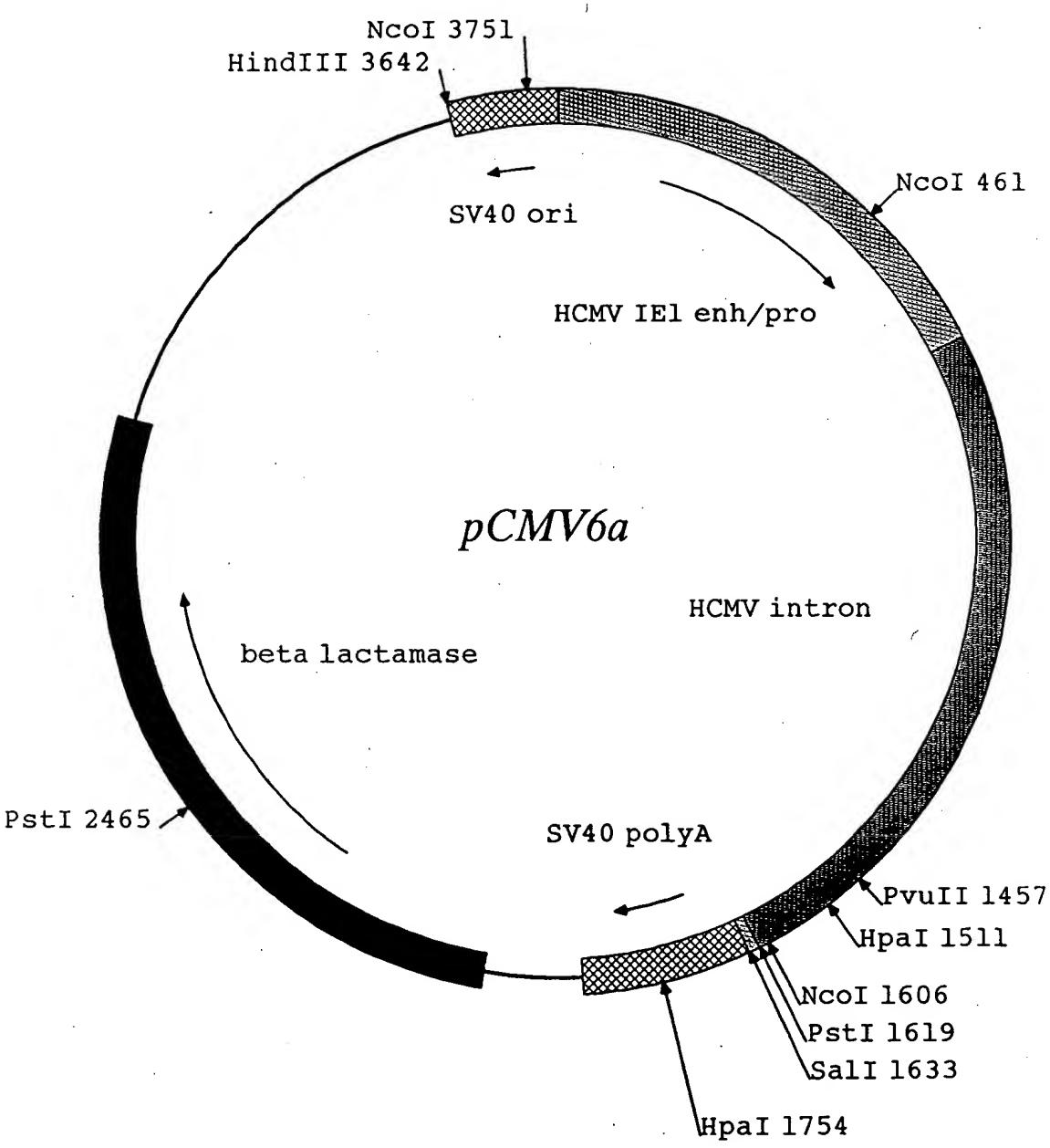

NcoI 3751
HindIII 3642
NcoI 461
SV40 ori
HCMV IE1 enh/pro
pCMV6a
HCMV intron
beta lactamase
PstI 2465
SV40 polyA
PvuII 1457
HpaI 1511
NcoI 1606
PstI 1619
SalI 1633
HpaI 1754

FIGURE 29

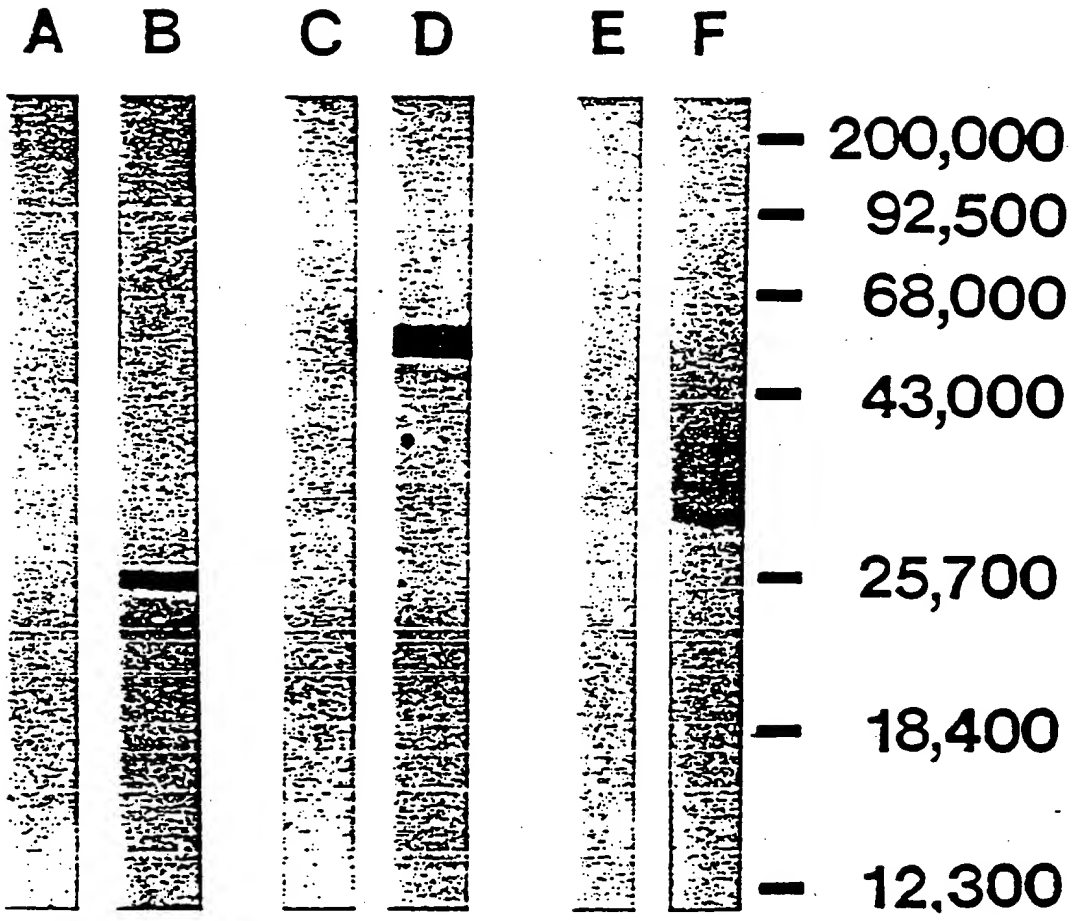
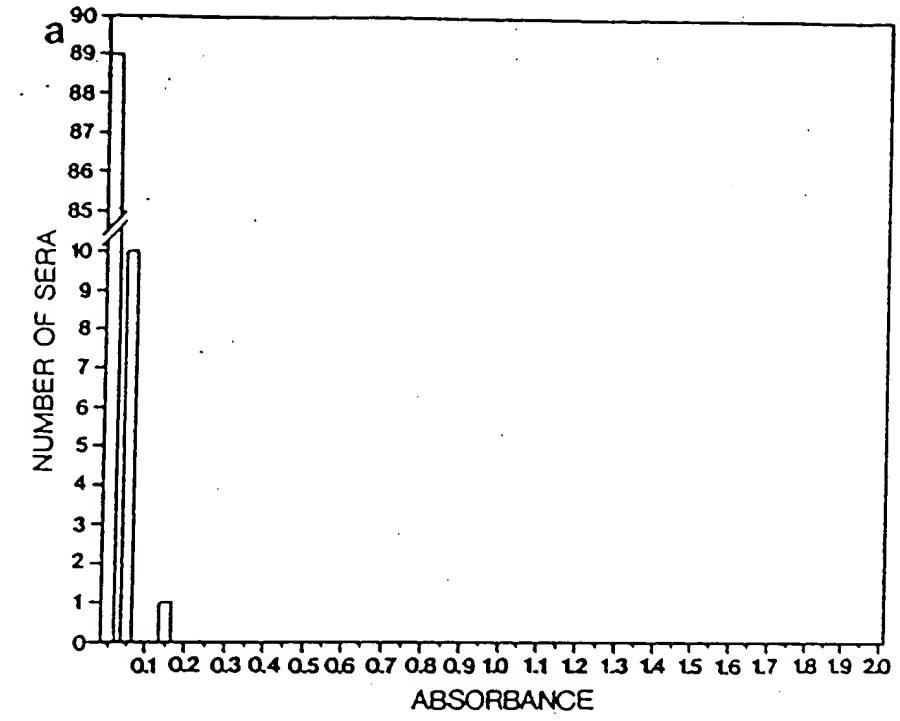


FIGURE 30


a
NUMBER OF SERA
0 1 2 3 4 5 6 7 8 9 10 85 86 87 88 89 90
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0
ABSORBANCE

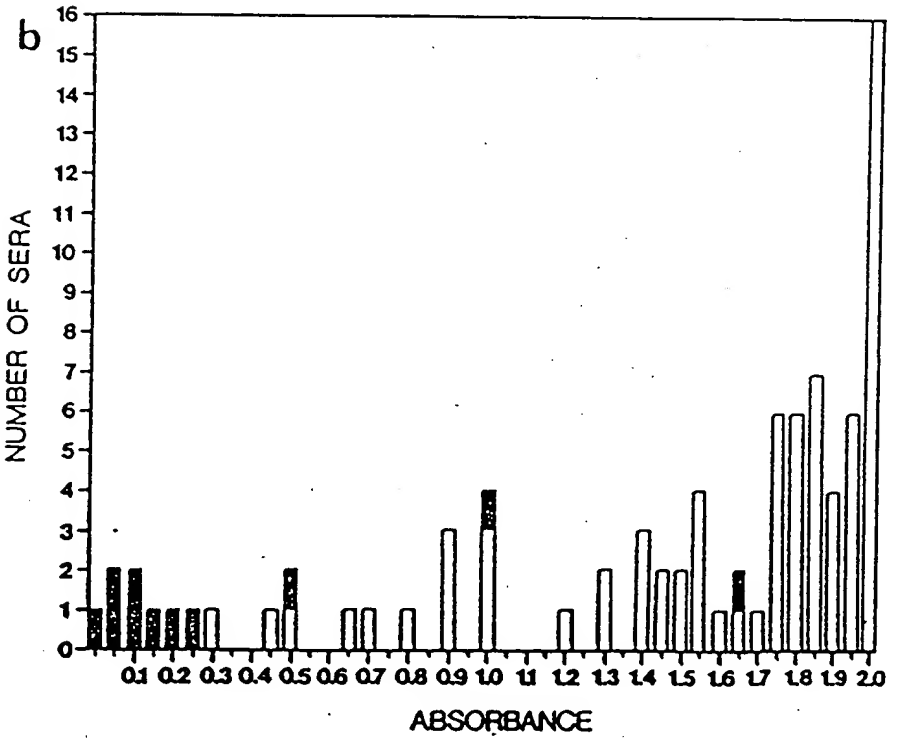

b
NUMBER OF SERA
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0
ABSORBANCE

FIGURE 31

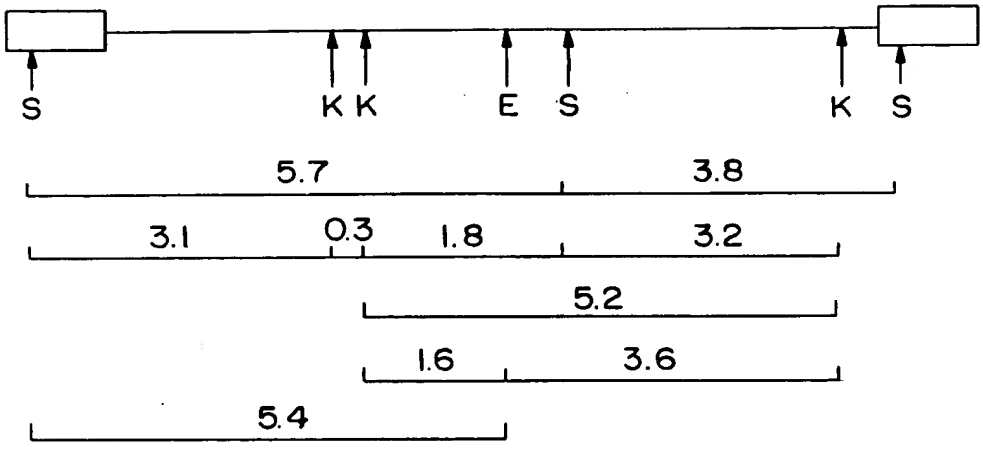


FIG. 1

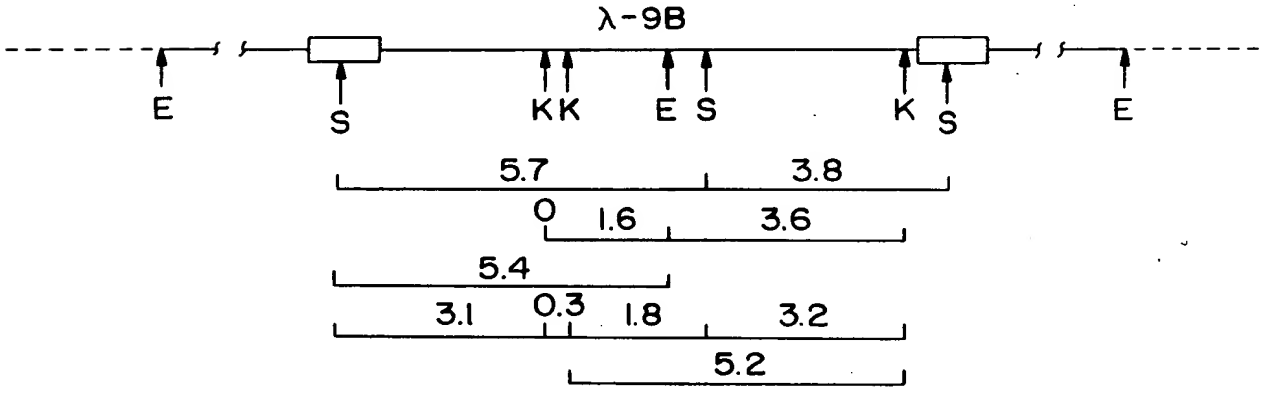


FIG. 2

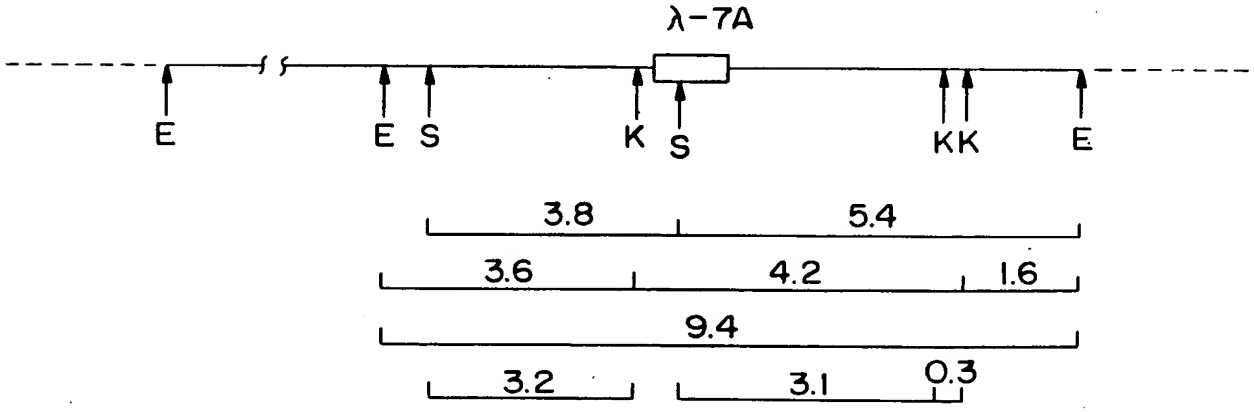


FIG. 3

Argument Map in DNA Strand ssarv2
from the '/v/lib/6mers'file.
Translation shown at open reading frames.

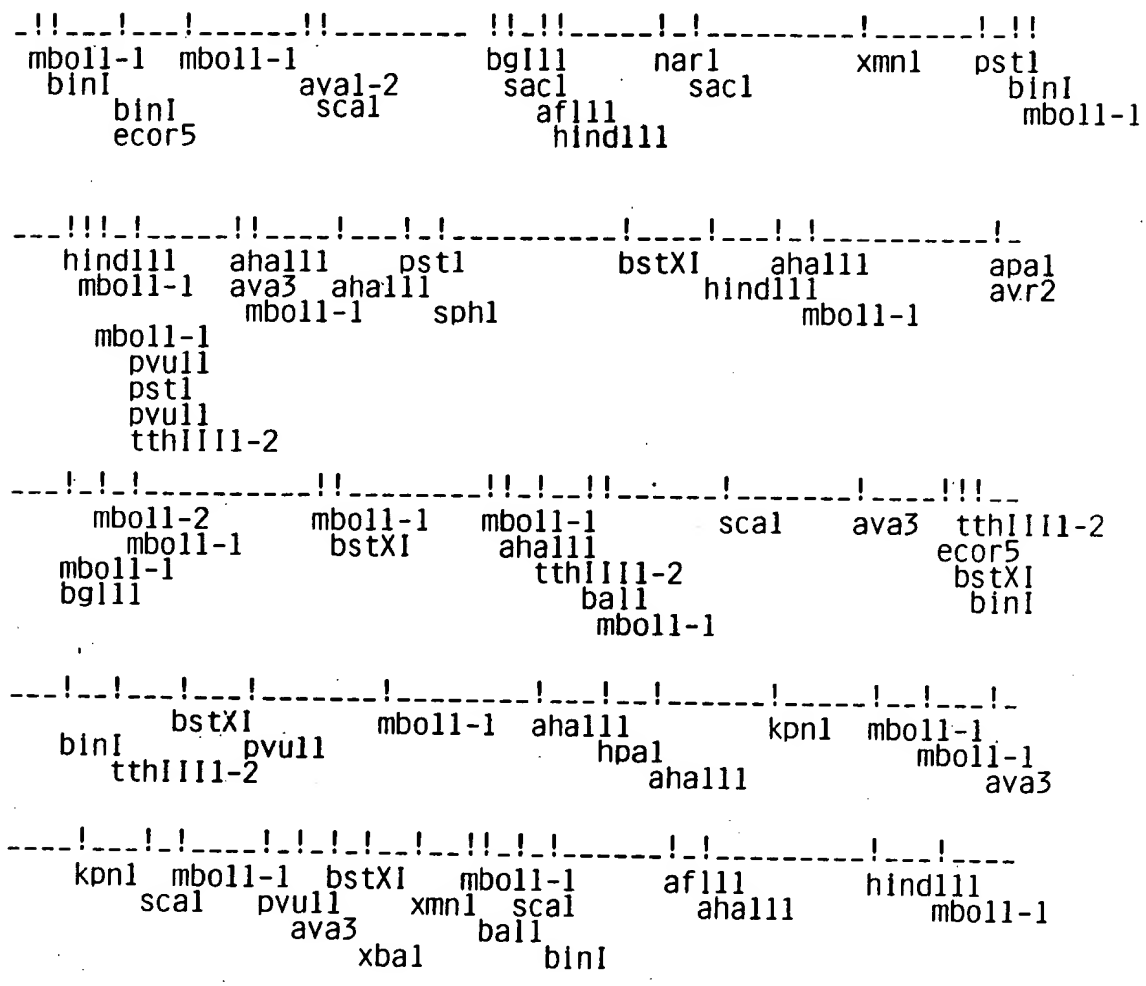


FIG. 4A

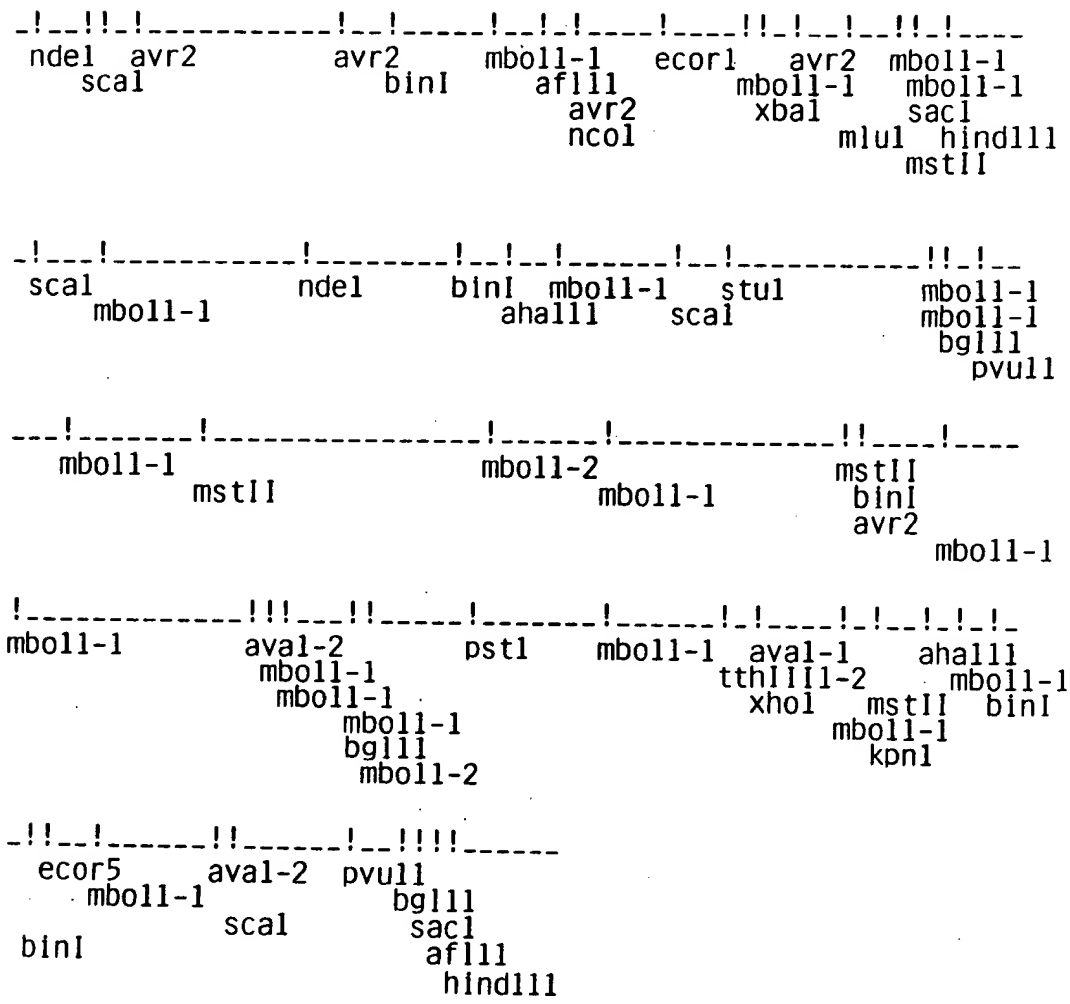

ndeI avrII scaI
avrII binI
mboII-1 aflII avrII ncoI
ecoRI avrII mboII-1 xbaI
mboII-1 mboII-1 sacI mluI hindIII mstII
scaI mboII-1
ndeI
binI mboII-1 ahaIII
stuI scaI
mboII-1 mboII-1 bglII pvuII
mboII-1 mstII
mboII-2 mboII-1
mstII binI avrII mboII-1
mboII-1 aval-2 mboII-1 mboII-1 mboII-1 bglII mboII-2
pstI
mboII-1 aval-1 tthIII-2 xhoI
ahaIII mboII-1 binI mstII mboII-1 kpnI
ecoR5 mboII-1 binI
aval-2 scaI
pvuII bglII sacI aflII hindIII

FIG. 4B

```
  1 CTGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACAC
    GACCTTCCCGATTAAACCAGGGTTTCTTCTGTTCTCTAGGAACTAGACACCTAGATGGTGTG

    26 mboll, 50 binI,

 63 ACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACT
    TGTTCCGATGAAGGGACTAACCGTCTTAATGTGTGGTCCCGGTCCCTAGTCTATAGGTGA

    107 binI, 113 ecor5,

123 GACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAA
    CTGGAAACCTACCACGAAGTTCGATCATGGTCAACTCGGTCTCTTCCATCTTCTCCGGTT

    172 mboll,

183 TGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGA
    ACTTCCTCTCTTGTTGTCGAACAATGTGGGATACTCGGACGTACCCTACCTCCTGCGCCT

243 GAAAGAAGTGTTAGTGTGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGA
    CTTTCTTCACAATCACACCTCCAAACTGTCGTTTGATCGTAAAGTAGTGTACCGGGCTCT

    296 aval,

303 GCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGC
    CGACGTAGGCCTCATGATGTTTCTGACGACTGTAGCTCGAAAGATGTTCCCTGAAAGGCG

    314 scal,

363 TGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGC
    ACCCCTGAAAGGTCCCTCCGCACCGGACCCGCCCTGACCCCTCACCGCAGGGAGTCTACG

423 TGCATATAAGCAGACTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAG
    ACGTATATTCGTCTGACGAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTC

    474 bgl11,

483 CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
    GGACCCTCGAGAGACCGATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACGGAA

    488 sac1, 518 afl11, 532 hindlll,

543 GAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
    CTCACGAAGTTCATCACACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAGT

603 GACCCTTTTAGTCAGTGTGGAAAAATCTCTAGCAGTGGCGCCCGAACAGGGACGCGAAAG
    CTGGGAAAATCAGTCACACCTTTTTAGAGATCGTCACCGCGGGCTTGTCCCTGCGCTTTC

    639 nar1,

663 CGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACAG
    GCTTTCATCTTGGTCTCCTCGAGAGAGCTGCGTCCTGAGCCGAACGACTTCGCGCGTGTC

    680 sac1,

723 CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAG
    GTTCTCCGCTCCCCGCCGCTGACCACTCATGCGGTTAAAAACTGATCGCCTCCGATCTTC

                MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGlu GAG
783 GAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA
    CTCTCTCTCTACCCACGCTCTCGCAGCCATAATTCGCCCCCTCTTAATCTATTTACCCTT
```

FIG. 4C

```
                LysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAla
843             AAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAACATATAGTATGGGCA
                TTTTAAGCCAATTCCGGTCCCCCTTTCTTTTTTATATTCAATTTTGTATATCATACCCGT

                SerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCys
903             AGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGCTGC
                TCGTCCCTCGATCTTGCTAAGCGTCAGTTAGGACCGGACAATCTTTGTAGTCTTCCGACG

                959 pstl,

                ArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSer
963             AGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAAGAACTTAGATCA
                TCTGTTTATAACCCTGTCGATGTCGGTAGGGAAGTCTGTCCTAGTCTTCTTGAATCTAGT

                1002 binI, 1008 mboll,

                LeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThr
1023            TTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACC
                AATATATTATGTCATCGTTGGGAGATAACACATGTAGTTTCCTATCTACATTTTCTGTGG

                LysGluAlaLeuGluLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGln
1083            AAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCAA
                TTCCTTCGAAATCTCTTCTATCTCCTTCTCGTTTTGTTTTCATTCTTTTTCCGTGTCGTT

                1087 hindlll, 1097 mboll, 1107 mboll,                    p25

                AlaAlaAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleVal
1143            GCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAAAATTACCCTATAGTG
                CGTCGTCGACGTCGACCGTGTCCTTTGTCGTCGGTCCAGTCGGTTTTAATGGGATATCAC

                1147 pvull, 1150 pstl, 1153 pvull, 1156 tthIII,

                GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
1203            CAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
                GTCTTGGATGTCCCCGTTTACCATGTAGTCCGGTATAGTGGATCTTGAAATTTACGTACC

                1250 ahalll, 1255 ava3,

                ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu
1263            GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
                CATTTTCATCATCTTCTTTTCCGAAAGTCGGGTCTTCATTATGGGTACAAAAGTCGTAAT

                1275 mboll,

                SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
1323            TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
                AGTCTTCCTCGGTGGGGTGTTCTAAATTTGTGGTACGATTTGTGTCACCCCCCTGTAGTT

                1346 ahalll,

                AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal
1383            GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
                CGTCGGTACGTTTACAATTTTCTCTGATAGTTACTCCTTCGACGTCTTACCCTATCTCAC

                1423 pstl,

                HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
1443            CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC
                GTAGGTCACGTACGTCCCGGATAACGTGGTCCGGTTTACTCTCTTGGTTCCCCTTCACTG

                1451 sphl,
```

FIG. 4D

```
                IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro
1503            ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
                TATCGTCCTTGATGATCATGGGAAGTCCTTGTTTATCCTACCTACTGTTTATTAGGTGGA

                IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1563            ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA
                TAGGGTCATCCTCTTTAGATATTTTCTACCTATTAGGACCCTAATTTATTTTATCATTCT

                MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp
1623            ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT
                TACATATCGGGATGGTCGTAAGACCTGTATTCTGTTCCTGGTTTCCTTGGGAAATCTCTA

                1636 bstXI,

                TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1683            TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT
                ATACATCTGGCCAAGATATTTTGAGATTCTCGGCTTGTTCGAAGTGTCCTACATTTTTTA

                1720 hindlll,

                TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys
1743            TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA
                ACCTACTGTCTTTGGAACAACCAGGTTTTACGTTTGGGTCTAACATTCTGATAAAATTTT

                1796 ahalll,

                AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1803            GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA
                CGTAACCCTGGTCGTCGATGTGATCTTCTTTACTACTGTCGTACAGTCCCTCACCCCCCT

                1827 mboll,

                ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn
1863            CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC
                GGGCCGGTATTTCGTTCTCAAAACCGACTTCGGTACTCGGTTCATTGTTTAGGTCGATTG

                p18
                IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1923            ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT
                TATTACTACGTCTCTCCGTTAAAATCCTTGGTTTCTTTCTGACAATTCACAAAGTTAACA

                GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg
1983            GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA
                CCGTTTCTTCCCGTGTATCGGTTTTTAACGTCCCGGGGATCCTTTTTCCCGACAACCTCT

                2014 apal, 2019 avr2,

                CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
2043            TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG
                ACACCTTCCCTTCCTGTGGTTTACTTTCTAACGTGACTCTCTGTCCGATTAAAAAATCCC

                2102 mboll,

                LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro
2103            AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA
                TTCTAGACCGGAAGGATGTTCCCTTCCGGTCCCTTAAAAGAAGTCTCGTCTGGTCTCGGT

                2104 bglll, 2141 mboll,
```

FIG. 4E

```
                ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
2163            ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG
                TGTCGGGGTGGTCTTCTCTCGAAGTCCAAACCCCTCCTCTTTTGTTGAGGGAGAGTCTTC

                2175 mboII,

                GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn
2223            CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC
                GTCCTCGGCTATCTGTTCCTTGACATAGGAAATTGAAGGGAGTCTAGTGAGAAACCGTTG

                AspProSerSerGlnOC
2283            GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGA
                CTGGGGAGCAGTGTTATTCCTATCCCCCCGTTGATTTCCTTCGAGATAATCTATGTCCT

                                        MetAsnLeuProGlyLysTrpLysProLysMetIle
2342            GCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATA
                CGTCTACTATGTCATAATCTTCTTTACTTAAACGGTCCTTTTACCTTTGGTTTTTACTAT

                2360 mboII, 2375 bstXI,

                GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCys
2402            GGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGT
                CCCCCTTAACCTCCAAAATAGTTTCATTCTGTCATGCTAGTCTATGGACATCTTTAGACA

                GlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg
2462            GGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGA
                CCTGTATTTCGATATCCATGTCATAATCATCCTGGATGTGGACAGTTGTATTAACCTTCT

                2517 mboII,

                AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal
2522            AATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTA
                TTAGACAACTGAGTCTAACCAACATGAAATTTAAAGGGGTAATCAGGATAACTTTGACAT

                2548 ahaIII, 2577 tthIIII,

                ProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu
2582            CCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAA
                GGTCATTTTAATTTCGGTCCTTACCTACCGGGTTTTCAATTCGTTACCGGTAACTGTCTT

                2627 ball, 2639 mboII,

                GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSer
2642            GAAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCA
                CTTTTTTATTTTCGTAATCATCTCTATACATGTCTTTACCTTTTCCTTCCCTTTTAAAGT

                LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer
2702            AAAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGT
                TTTTAACCCGGACTTTTAGGTATGTTATGAGGTCATAAACGATATTTCTTTTTTCTGTCA

                2759 scaI,

                ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp
2762            ACTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGG
                TGATTTACCTCTTTTGATCATCTAAAGTCTCTTGAATTATTTTCTTGAGTTCTGAAGACC

                GluValGlnLeuGlyIleProHisProGlnGlyOC
2822            GAAGTTCAGTTAGGAATACCACACCCGCAGGGTTAAAAAAAGAAAAAATCAGTAACAGTA
                CTTCAAGTCAATCCTTATGGTGTGGGCGTCCCAATTTTTTTCTTTTTTAGTCATTGTCAT
```

FIG. 4F

```
2882 TTGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG
     AACCTACACCCACTACGTATGAAAAGTCAAGGGAATCTATTTCTGAAATCTTTCATATGAC

     2895 ava3,

                                        MetArgHisGlnGlyLeuAspIleSerThrMetTrp
2943 CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGG
     GTAAATGGTATGGATCATATTTGTTACTCTGTGGTCCCTAATCTATAGTCATGTTACACC

     2985 ecor5,

     LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeu
3003 CTGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTA
     GACGGTGTCCCTACCTTTCCTAGTGGTCGTTATAAGGTTTCATCGTACTGTTTTTAGAAT

     3003 tthIII1, 3006 bstXI, 3021 binI,

     GluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr
3063 GAGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTAT
     CTCGGAAAATCTTTTGTCTTAGGTCTGTATCAATAGATAGTTATGTACCTACTAAACATA

     ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHis
3123 GTAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCAT
     CATCCTAGACTGAATCTTTATCCCGTCGTATCTTGTTTTTATCTCCTTGACTCTGTCGTA

     3126 binI, 3171 tthIII1,

     LeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu
3183 CTGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTT
     GACAACTCCACCCCTAAATGGTGTGGTCTGTTTTTTGTAGTCTTTCTTGGAGGTAAGGAA

     3234 bstXI,

     TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGlu
3243 TGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAA
     ACCTACCCAATACTTGAGGTAGGACTATTTACCTGTCATGTCGGATATTACGACGGTCTT

     LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer
3303 AAAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGT
     TTTCTGTCGACCTGACAGTTACTGTATGTCTTCAATCACCCTTTTAACTTAACCCGTTCA

     3308 pvuII,

     GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAla
3363 CAGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCA
     GTCTAAATACGTCCCTAATTTCATTTCGTCAATACATTTGAGGAATCTCCTTGGTTTCGT

     LeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu
3423 CTAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAG
     GATTGTCTTCATTATGGTGATTGTCTTCTTCGTCTCGATCTTGACCGTCTTTTGTCCCTC

     3447 mboII,

     IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGlu
3483 ATTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAA
     TAAGATTTTCTTGGTCATGTACTTCATATAATACTGGGTAGTTTTCTGAATCATCGTCTT

     IleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn
3543 ATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAAT
     TATGTCTTCGTCCCCGTTCCGGTTACCTGTATAGTTTAAATAGTTCTCGGTAAATTTTTA

     3594 ahaIII,
```

POL

FIG. 4G

```
                LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeu
3603            CTGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTA
                GACTTTTGTCCTTTCATACGTTCCTACTCCCCACGGGTGTGATTACTACATTTTGTCAAT

                3659 hpa1,

                ThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLysIleProLys
3663            ACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAA
                TGTCTCCGTCACGTTTTTCATAGGTGTCTTTCGTATCATTATACCCCTTTCTAAGGATTT

                PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAla
3723            TTTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCT
                AAATTTGATGGGTATGTTTTCCTTTGTACCCTTCGTACCACCTACCTCATAACCGTTCGA

                3723 aha111,

                ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln
3783            ACCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAG
                TGGACCTAAGGACTCACCCTCAAACAGTTATGGGGAGGGAATCACTTTAATACCATGGTC

                3835 kpn1,

                LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg
3843            TTAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGG
                AATCTCTTTCTTGGGTATCATCCTCGTCTTTGAAAGATACATCTACCCCGTCGATTATCC

                GluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer
3903            GAGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCC
                CTCTGATTTAATCCTTTTCGTCCTATACAATGACTGTCTCCTTCTGTTTTTCAACAGAGG

                3943 mbo11,

                IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAsp
3963            ATAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGAT
                TATCGACTGTGTTGTTTAGTCTTCTGACTTAATGTTCGTTAAGTAGATCGAAACGTCCTA

                3983 mbo11,

                SerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla
4023            TCGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCA
                AGCCCTAATCTTCATTTGTATCATTGTCTGAGTGTTATACGTAATCCTTAGTAAGTTCGT

                4060 ava3,

                GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLys
4083            CAACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAG
                GTTGGTCTATTCTCACTTAGTCTCAATCAGTCAGTTTATTATCTCGTCAATTATTTTTTC

                GluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal
4143            GAAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTA
                CTTTTCCAGATGGACCGTACCCATGGTCGTGTGTTTCCTTAACCTCCTTTACTTGTTCAT

                4163 kpn1,

                AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAla
4203            GATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCC
                CTATTTAATCAGTCACGACCTTAGTCCTTTCATGATAAAAACTTACCTTATCTATTCCGG

                4232 sca1,
```

FIG. 4H

```
                GlnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu
4263            CAAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTG
                GTTCTTCTTGTACTCTTTATAGTGTCATTAACCTCTCGTTACCGATCACTAAAATTGGAC
                     ^
                4266 mbo11,

                ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGlu
4323            CCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAA
                GGTGGACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTCGATTTTCCTCTT
                                            ^
                4352 pvu11,

                AlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu
4383            GCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTA
                CGGTACGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTAACATGTGTAGAT
                     ^                  ^                       ^
                4386 ava3, 4410 bstXI, 4439 xba1,

                GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluVal
4443            GAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTT
                CTTCCTTTTTAATAGGACCATCGTCAAGTACATCGGTCACCTATATATCTTCGTCTTCAA
                                                              ^
                4497 xmn1,

                IleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp
4503            ATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGG
                TAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATAAAAGAGAATTTTAATCGTCCTTCTACC
                                                        ^     ^
                4555 mbo11, 4560 ba11,

                ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAla
4563            CCAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCC
                GGTCATTTTTGTTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATGATGCCAATTCCGG
                                                        ^
                4605 sca1,

                AlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln
4623            GCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAA
                CGGACAACCACCCGTCCCTAGTTCGTCCTTAAACCGTAAGGGATGTTAGGGGTTTCAGTT
                                  ^
                4639 bin1,

                GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGln
4683            GGAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAG
                CCTCATCATCTTAGATACTTATTACTTAATTTCTTTTAATATCCTGTCCATTCTCTAGTC

                AlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys
4743            GCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAA
                CGACTTGTGGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTAAAATTTTCTTTT
                      ^                                       ^
                4752 alf11, 4791 aha111,

                                rSerAlaGlyGluArgIleValAspIleIleAlaThrAspIle
48(                             CAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATA
                                GTCACGTCCCCTTTCTTATCATCTGTATTATCGTTGTCTGTAT
                qfp11
                                lLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg
48E                             AAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGG
                                TTCGTTTAATGTTTTTAAGTTTTAAAAGCCCAAATAATGTCC
```

FIG. 4 I

```
                AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAla
4923            GACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCA
                CTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGTTTCGAAGAGACCTTTCCACTTCCCCGT

                4956 hindlll,

                ValValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle
4983            GTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATT
                CATCATTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTTCGTTTTTAGTAA

                5023 mboll,

                      MetGluAsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIle
                ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAsp
5043            AGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGAT
                TCCCTAATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCTGTCCTACTCCTA

                 ArgTreTrpLysSerLeuValLysHisHisMetTyrIleSerLysLysAlaLysGlyTrp
                AM
5103            TAGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG
                ATCTTGTACCTTTTCAAATCATTTTGTGGTATACATATAAAGTTTCTTTCGATTTCCTACC

                5131 ndel,

                   PheTyrArgHisHisTyrGluSerThrHisProArgValSerSerGluValHisIle
5163               TTTTATAGACATCACTATGAAAGTACTCATCCAAGAGTAAGTTCAGAAGTACACATC
                   AAAATATCTGTAGTGATACTTTCATGAGTAGGTTCTCATTCAAGTCTTCATGTGTAG

                5185 scal,

                ProLeuGlyAspAlaLysLeuValIleThrThrTyrTrpGlyLeuHisThrGlyGluArg
5221            CCCCTAGGGGATGCTAAATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGA
                GGGGATCCCCTACGATTTAACCATTATTGTTGTATAACCCCAGACGTATGTCCTCTTTCT

                5223 avr2,

                GluTrpHisLeuGlyGlnGlyValAlaIleGluTrpArgLysLysLysTyrSerThrGln
5281            GAATGGCATTTGGGCCAGGGAGTCGCCATAGAATGGAGGAAAAAGAAATATAGCACACAA
                CTTACCGTAAACCCGGTCCCTCAGCGGTATCTTACCTCCTTTTTCTTTATATCGTGTGTT

                ValAspProGlyLeuAlaAspGlnLeuIleHisLeuHisTyrPheAspCysPheSerGlu
5341            GTAGACCCTGGCCTAGCAGACCAACTAATTCATCTGCATTATTTTGATTGTTTTTCAGAA
                CATCTGGGACCGGATCGTCTGGTTGATTAAGTAGACGTAATAAAACTAACAAAAAGTCTT

                SerAlaIleLysAsnAlaIleLeuGlyTyrArgValSerProArgCysGluTyrGlnAla
5401            TCTGCTATAAAAAATGCCATATTAGGATATAGAGTTAGTCCTAGGTGTGAATATCAAGCA
                AGACGATATTTTTTACGGTATAATCCTATATCTCAATCAGGATCCACACTTATAGTTCGT

                5440 avr2,

                GlyHisAsnLysValGlySerLeuGlnTyrLeuAlaLeuAlaAlaLeuIleThrProLys
5461            GGACATAACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAA
                CCTGTATTGTTCCATCCTAGAGATGTTATGAACCGTGATCGTCGTAATTATTGTGGTTTT

                5476 binl,

                LysThrLysProProLeuProSerValLysLysLeuThrGluAspArgTrpAsnLysPro
5521            AAGACAAAGCCACCTTTGCCTAGTGTTAAGAAACTGACAGAGGATAGATGGAACAAGCCC
                TTCTGTTTCGGTGGAAACGGATCACAATTCTTTGACTGTCTCCTATCTACCTTGTTCGGG
```

FIG. 4J

```
                GlnLysThrLysGlyHisArgGlySerHisThrMetAsnGlyHisAM
5581 CAGAAGACCAAGGGCCACAGAGGGAGCCATACAATGAATGGACACTAGAGCTTTTAGAGG
     GTCTTCTGGTTCCCGGTGTCTCCCTCGGTATGTTACTTACCTGTGATCTCGAAAATCTCC

     5583 mboII,

5641 AGCTTAAGAGAGAAGCTGTTAGACATTTTCCTAGGCCATGGCTCCATAGCTTAGGACAAT
     TCGAATTCTCTCTTCGACAATCTGTAAAAGGATCCGGTACCGAGGTATCGAATCCTGTTA

     5643 afIII, 5670 avr2, 5676 ncoI,

5701 ATATCTATGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGC
     TATAGATACTTTGAATACCCCTATGAACCCGTCCTCACCTTCGGTATTATTCTTAAGACG

     5752 ecorI,

5761 AACAACTGCTGTTTATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTATTC
     TTGTTGACGACAAATAAGTAAAGTCTTAACCCACAGTTGTATCGTCTTATCCGTAATAAG

5821 AACAGAGGAGAGCAAGAAGAAATGGAGCCAGTAGATCCTAATCTAGAGCCCTGGAAGCAT
     TTGTCTCCTCTCGTTCTTCTTTACCTCGGTCATCTAGGATTAGATCTCGGGACCTTCGTA

     5836 mboII, 5862 xbaI,

5881 CCAGGAAGTCAGCCTAGGACTGCTTGTAACAATTGCTATTGTAAAAAGTGTTGCTTTCAT
     GGTCCTTCAGTCGGATCCTGACGAACATTGTTAACGATAACATTTTTCACAACGAAAGTA

     5893 avr2,

5941 TGCTACGCGTGTTTCACAAGAAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGA
     ACGATGCGCACAAAGTGTTCTTTTCCGAATCCGTAGAGGATACCGTCCTTCTTCGCCTCT

     5945 mluI, 5988 mboII,

6001 CAGCGACGAAGAGCTCCTCAGGACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAA
     GTCGCTGCTTCTCGAGGAGTCCTGTCAGTCTGAGTAGTTCGAAGAGATAGTTTCGTCATT

     6008 mboII, 6011 sacI, 6016 mstII, 6038 hindIII,

6061 GTAGTAAATGTAATGCAATCTTTACAAATATTAGCAATAGTATCATTAGTAGTAGTAGCA
     CATCATTTACATTACGTTAGAAATGTTTATAATCGTTATCATAGTAATCATCATCATCGT

6121 ATAATAGCAATAGTTGTGTGGACCATAGTACTCATAGAATATAGGAAAATATTAAGACAA
     TATTATCGTTATCAACACACCTGGTATCATGAGTATCTTATATCCTTTTATAATTCTGTT

     6147 scaI,

                                                          MetLys  ENV
6181 AGAAAATAGACAGATTAATTGATAGAATAAGAGAAAAAGCAGAAGACAGTGGCAATGAAA
     TCTTTTATCTGTCTAATTAACTATCTTATTCTCTTTTTCGTCTTCTGTCACCGTTACTTT

     6222 mboII,

     ValLysGlyThrArgArgAsnTyrGlnHisLeuTrpArgTrpGlyThrLeuLeuLeuGly
6241 GTGAAGGGGACCAGGAGGAATTATCAGCACTTGTGGAGATGGGGCACCTTGCTCCTTGGG
     CACTTCCCCTGGTCCTCCTTAATAGTCGTGAACACCTCTACCCCGTGGAACGAGGAACCC

     MetLeuMetIleCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValPro
6301 ATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCT
     TACAACTACTAGACATCACGATGTCTTTTTAACACCCAGTGTCAAATAATACCTCATGGA
```

FIG. 4K

```
                ValTrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThr
6361            GTGTGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACA
                CACACCTTTCTTCGTTGATGGTGAGATAAAACACGTAGTCTACGATCTCGTATACTATGT

                6410 ndeI,

                GluValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGlu
6421            GAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAA
                CTCCATGTATTACAAACCCGGTGTGTACGGACACATGGGTGTCTGGGGTTGGGTGTTCTT

                ValValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGln
6481            GTAGTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAG
                CATCATAACCCTTTACACTGTCTTTTAAAATTGTACACCTTTTTATTGTACCATCTTGTC

                MetGlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThr
6541            ATGCAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACC
                TACGTCCTCCTATATTAGTCAAATACCCTAGTTTCGGATTTCGGTACACATTTTAATTGG

                6567 binI,

                ProLeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSer
6601            CCACTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGT
                GGTGAGACACAATGAAATTTAACGTGACTAAACCCCTTCCGATGATTATGGTTATCATCA

                6615 ahaIII,

                AsnTrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSer
6661            AATTGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGC
                TTAACCTTTCTTCTTTATTTTCCTCTTTATTTTTTGACGAGAAAGTTATAGTGGTGTTCG

                6670 mboII,

                IleArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIle
6721            ATAAGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATA
                TATTCTCTATTCTAAGTCTTTCTTTTACGTGAAAAAGCATTGGAACTACATCATGGTTAT

                AspAsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSer
6781            GATAATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCA
                CTATTACGATCATGATGATGGTTGATATGGTTGATATCCAACTATGTAACATTGTCTAGT

                6790 scaI,

                ValIleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThr
6841            GTCATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACC
                CAGTAATGTGTCCGGACAGGTTTCCATAGTAAACTCGGTTAAGGGTATGTAATAACATGG

                6851 stuI,

                ProAlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCys
6901            CCGGCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGT
                GGCCGACCAAAACGCTAAGATTTCACATTATTATTTTGCAAGTTACCTTTTCCTGGTACA

                ThrAsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeu
6961            ACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTG
                TGTTTACAGTCGTGTCATGTTACATGTGTACCTTAATCCGGTTATCACAGTTGAGTTGAC

                LeuLeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsn
7021            CTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAAC
                GACAATTTACCGTCAGATCGTCTTCTTCTCCATCATTAATCTAGACTGTTAAAGTGCTTG

                7042 mboII, 7045 mboII, 7060 bglII,
```

FIG. 4L

```
                AsnAlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgPro
7081            AATGCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCC
                TTACGATTTTGGTATTATCATGTCGACTTACTTAGACATCGTTAATTGACATGTTCTGGG

                7102 pvuII,

                AsnAsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGly
7141            AACAACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGA
                TTGTTGTTATGTTCTTTTTCATAGATATATCCTGGTCCCTCTCGTAAAGTATGTTGTCCT

                7199 mboII,

                ArgIleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsn
7201            AGAATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAAC
                TCTTATTATCCTCTATATTCTTTTCGTGTAACATTGTAATCATCTCGTGTTACCTTATTG

                ThrLeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleVal
7261            ACTTTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTC
                TGAAATCTTGTCTATCAATTTTTTAATTCTCTTGTCAAACCCTTATTATTTTGTTATCAG

                PheAsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGly
7321            TTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGG
                AAATTAGTTAGGAGTCCTCCCCTGGGTCTTTAACATTACGTGTCAAAATTAACATCTCCC

                7331 mstII,

                GluPhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThr
7381            GAATTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACT
                CTTAAAAAGATGACATTATGTTGTGTTGACAAATTATTATGTACCTCCAATTTAGTGTGA

                GluGlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsn
7441            GAAGGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAAC
                CTTCCTTGATTTCCTTTACTGTGTTAGTATGAGGGTACATCTTATTTTGTTTAATATTTG

                MetTrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCys
7501            ATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGT
                TACACCGTCCTTCATCCTTTTCGTTACATACGGGGAGGGTAACCTCCTGTTTAATCAACA

                SerSerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAsp
7561            TCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGAC
                AGTAGTTTATAATGTCCCGACGATAATTGTTCTCTACCACCATGTTTACATTGATTACTG

                ThrGluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyr
7621            ACCGAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATAT
                TGGCTCCAGAAGTCTGGACCTCCTCCTCTATACTCCCTGTTAACCTCTTCACTTAATATA

                7628 mboII,

                LysTyrLysValIleLysIleGluProLeuGlyIleAlaProThrLysAlaLysArgArg
7681            AAATATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACCAAGGCAAAGAGAAGA
                TTTATATTTCATTATTTTTAACTTGGTAATCCTTATCGTGGGTGGTTCCGTTTCTCTTCT

                7736 mboII,

                ValValGlnArgGluLysArgAlaValGlyIleValGlyAlaMetPheLeuGlyPheLeu
7741            GTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGTAGGAGCTATGTTCCTTGGGTTCTTG
                CACCACGTCTCTCTTTTTTCTCGTCACCCTTATCATCCTCGATACAAGGAACCCAAGAAC

                GlyAlaAlaGlySerThrMetGlyAlaValSerLeuThrLeuThrValGlnAlaArgGln
7801            GGAGCAGCAGGAAGCACTATGGGCGCAGTGTCATTGACGCTGACGGTACAGGCCAGACAA
                CCTCGTCGTCCTTCGTGATACCCGCGTCACAGTAACTGCGACTGCCATGTCCGGTCTGTT
```

FIG. 4M

```
      LeuLeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGln
7861  TTATTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAA
      AATAACAGACCATATCACGTTGTCGTCTTGTTAAACGACTCCCGATAACTCCGCGTTGTT

      HisLeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaVal
7921  CATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTG
      GTAGACAACGTTGAGTGTCAGACCCCGTAGTTCGTCGAGGTCCGTTCTCAGGACCGACAC

      GluArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIle
7981  GAAAGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATT
      CTTTCTATGGATTCCCTAGTTGTCGAGGATCCCTAAACCCCAACGAGACCTTTTGAGTAA

      7989 mstII, 7995 binI, 8007 avr2,

      CysThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrp
8041  TGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAAGACATTTGG
      ACGTGGTGACGACACGGAACCTTACGATCAACCTCATTATTTAGAGACCTTCTGTAAACC

      8089 mboII,

      AspAsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyr
8101  GATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAACACAATATAC
      CTATTGTACTGGACCTACGTCACCCTTTCTCTTTAACTGTTAATGTGTTTGTGTTATATG

      ThrLeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeu
8161  ACCTTACTTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTG
      TGGAATGAACTTCTTAGCGTCTTGGTTGTTCTTTTCTTACTTGTTCTTAATAATCTTAAC

      8170 mboII,

      AspLysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpLeuTrpTyrIleLys
8221  GATAAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAAG
      CTATTCACCCGTTCAAACACCTTAACCAAATCGTATTGTTTGACCGACACCATATATTTC

      IlePheIleMetIleValGlyGlyLeuValGlyLeuArgIleValPheAlaValLeuSer
8281  ATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTCT
      TATAAGTATTACTATCATCCTCCGAACCATCCAAATTCTTATCAAAAACGACACGAAAGA

      IleValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrArgLeuProVal
8341  ATAGTGAATAGAGTTAGGCAGGGATACTCACCATTGTCATTTCAGACCCGCCTCCCAGTC
      TATCACTTATCTCAATCCGTCCCTATGAGTGGTAACAGTAAAGTCTGGGCGGAGGGTCAG

      8400 aval,

      ProArgGlyProAspArgProAspGlyIleGluGluGluGlyGlyGluArgAspArgAsp
8401  CCGAGGGGACCCGACAGGCCCGACGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGAC
      GGCTCCCCTGGGCTGTCCGGGCTGCCTTAGCTTCTTCTTCCACCTCTCTCTCTGTCTCTG

      8431 mboII, 8434 mboII,

      ArgSerValArgLeuValAspGlyPheLeuAlaLeuIleTrpGluAspLeuArgSerLeu
8461  AGATCCGTTCGATTAGTGGATGGATTCTTAGCACTTATCTGGGAAGATCTGCGGAGCCTG
      TCTAGGCAAGCTAATCACCTACCTAAGAATCGTGAATAGACCCTTCTAGACGCCTCGGAC

      8503 mboII, 8505 bglII,

      CysLeuPheSerTyrArgArgLeuArgAspLeuLeuLeuIleAlaAlaArgThrValGlu
8521  TGCCTCTTCAGCTACCGCCGCTTGAGAGACTTACTCTTGATTGCAGCGAGGACTGTGGAA
      ACGGAGAAGTCGATGGCGGCGAACTCTCTGAATGAGAACTAACGTCGCTCCTGACACCTT

      8525 mboII,
```

FIG. 4N

```
                IleLeuGlyHisArgGlyTrpGluAlaLeuLysTyrTrpTrpSerLeuLeuGlnTyrTrp
8581  ATTCTGGGGCACAGGGGGTGGGAAGCCCTCAAATATTGGTGGAGTCTCCTGCAGTATTGG
      TAAGACCCCGTGTCCCCCACCCTTCGGGAGTTTATAACCACCTCAGAGGACGTCATAACC

      8629 pst1,

                IleGlnGluLeuLysAsnSerAlaValSerTrpLeuAsnAlaThrAlaIleAlaValThr
8641  ATTCAGGAACTAAAGAATAGTGCTGTTAGCTGGCTCAACGCCACAGCTATAGCAGTAACT
      TAAGTCCTTGATTTCTTATCACGACAATCGACCGAGTTGCGGTGTCGATATCGTCATTGA

                GluGlyThrAspArgValIleGluValAlaGlnArgAlaTyrArgAlaIleLeuHisIle
8701  GAGGGGACAGATAGGGTTATAGAAGTAGCACAAAGAGCTTATAGAGCTATTCTCCACATA
      CTCCCCTGTCTATCCCAATATCTTCATCGTGTTTCTCGAATATCTCGATAAGAGGTGTAT

                HisArgArgIleArgGlnGlyLeuGluArgLeuLeuLeuOC  MetGlyGlyLysTrpSer
8761  CATAGAAGAATTAGACAGGGCTTGGAAAGGCTTTTGCTATAAGATGGGTGGCAAGTGGTCA
      GTATCTTCTTAATCTGTCCCGAACCTTTCCGAAAACGATATTCTACCCACCGTTCACCAGT

      8765 mboll,

                LysArgSerMetGlyGlyTrpSerAlaIleArgGluArgMetArgArgAlaGluProArg
8822  AAACGTAGTATGGGTGGATGGTCTGCTATAAGGGAAAGAATGAGACGAGCTGAGCCACGA
      TTTGCATCATACCCACCTACCAGACGATATTCCCTTTCTTACTCTGCTCGACTCGGTGCT

                AlaGluProAlaAlaAspGlyValGlyAlaValSerArgAspLeuGluLysHisGlyAla
8882  GCTGAGCCAGCAGCAGATGGGGTGGGAGCAGTATCTCGAGACCTGGAAAAACATGGAGCA
      CGACTCGGTCGTCGTCTACCCCACCCTCGTCATAGAGCTCTGGACCTTTTTGTACCTCGT

      8883 tthIII1, 8916 aval xho1,

                IleThrSerSerAsnThrAlaAlaThrAsnAlaAspCysAlaTrpLeuGluAlaGlnGlu
8942  ATCACAAGTAGCAATACAGCAGCTACTAATGCTGATTGTGCCTGGCTAGAAGCACAAGAG
      TAGTGTTCATCGTTATGTCGTCGATGATTACGACTAACACGGACCGATCTTCGTGTTCTC

                GluGluGluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLys
9002  GAGGAAGAGGTGGGTTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTACAAG
      CTCCTTCTCCACCCAAAAGGTCAGTCTGGAGTCCATGGAAATTCTGGTTACTGAATGTTC

      9005 mboll, 9029 mstII, 9034 kpn1,

                AlaAlaLeuAspIleSerHisPheLeuLysGluLysGlyGlyLeuGluGlyLeuIleTrp
9062  GCAGCTTTAGATATTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTTGG
      CGTCGAAATCTATAATCGGTGAAAAATTTTCTTTTCCCCCCTGACCTTCCCGATTAAACC

      9085 aha111,

                SerGlnArgArgGlnGluIleLeuAspLeuTrpIleTyrHisThrGlnGlyTyrPhePro
9122  TCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCT
      AGGGTTTCTTCTGTTCTCTAGGAACTAGACACCTAGATGGTGTGTGTTCCGATGAAGGGA

      9129 mboll, 9153 binI,

                AspTrpGlnAsnTyrThrProGlyProGlyIleArgTyrProLeuThrPheGlyTrpCys
9182  GATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGC
      CTAACCGTCTTAATGTGTGGTCCCGGTCCCTAGTCTATAGGTGACTGGAAACCTACCACG

      9210 binI, 9216 ecor5,
```

FIG. 40

```
                PheLysLeuValProValGluProGluLysValGluGluAlaAsnGluGlyGluAsnAsn
9242  TTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAAC
      AAGTTCGATCATGGTCAACTCGGTCTCTTCCATCTTCTCCGGTTACTTCCTCTCTTGTTG

      9275 mboll,

                SerLeuLeuHisProMetSerLeuHisGlyMetGluAspAlaGluLysGluValLeuVal
9302  AGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGAGAAAGAAGTGTTAGTG
      TCGAACAATGTGGGATACTCGGACGTACCCTACCTCCTGCGCCTCTTTCTTCACAATCAC

                TrpArgPheAspSerLysLeuAlaPheHisHisMetAlaArgGluLeuHisProGluTyr
9362  TGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTAC
      ACCTCCAAACTGTCGTTTGATCGTAAAGTAGTGTACCGGGCTCTCGACGTAGGCCTCATG

      9399 aval, 9417 scal,

                TyrLysAspCysOP
9422  TACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGG
      ATGTTTCTGACGACTGTAGCTCGAAAGATGTTCCCTGAAAGGCGACCCCTGAAAGGTCCC

9482  AGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGCTGCATATAAGCAGCTG
      TCCGCACCGGACCCGCCCTGACCCCTCACCGCAGGGAGTCTACGACGTATATTCGTCGAC

      9536 pvull,

9542  CTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGC
      GAAAAACGGACATGACCCAGAGAGACCAATCTGGTCTAGACTCGGACCCTCGAGAGACCG

      9576 bglll, 9590 sacl,

9602  TAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTG
      ATTGATCCCTTGGGTGACGAATTCGGAGTTATTTCGAACGGAACTCACGAAGTTCATCAC

      9620 afllI, 9634 hindlll,

9662  TGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTG
      ACACGGGCAGACAACACACTGAGACCATTGATCTCTAGGGAGTCTGGGAAAATCAGTCAC

9722  TGGAAAAATCTCTAGCAG
      ACCTTTTTAGAGATCGTC
```

FIG. 4P

```
      U3→
-453 CTGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCAC
     ACACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCA
-333 CTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCC
     AATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCG
-214 GAGAAAGAAGTGTTAGTGTGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGA
     GAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCG
 -93 CTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATG
                                    ←U3 R→
     CTGCATATAAGCAGCTGCTTTTTGCCTGTACTG GGTCTCTCTGGTTAGACCAGATCTGAG
  28 CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
        ←R U5→
     GAGTGCTTCA AGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
                                   ←U5
 148 GACCCTTTTAGTCAGTGTGGAAAAATCTCTAGCAG TGGCGCCCGAACAGGGACGCGAAA
     GCGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACAG

 268 CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAG
           MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGlu    17
     GAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA
     LysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAla
 388 AAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAACATATAGTATGGGCA
     SerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCys    57
     AGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGCTGC
     ArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSer
 508 AGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAAGAACTTAGATCA
     LeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThr    97
     TTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACC
     LysGluAlaLeuGluLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGln
 628 AAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCAA
     AlaAlaAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleVal   137
     GCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAAAATTACCCTATAGTG
     GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
 748 CAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
     ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu   177
     GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
     SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
 868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
```

FIG. 5A

```
      AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal  217
      GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG

      HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
  988 CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC

      IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro  257
      ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT

      IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
 1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA

      MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp  297
      ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT

      TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
 1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT

      TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys  337
      TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA

      AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
 1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA

      ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn  377
      CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC

      IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
 1468 ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT

      GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg  417
      GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA

      CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
                                                          PhePheArgG
 1588 TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG

      LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro  457
      luAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAla  23
      AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA

      ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
     AsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSerLeuSerGluA
 1708 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG

      GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn  497
      laGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGln  63
      CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC

      AspProSerSerGlnOC
     ArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyA
 1828 GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAG

      laAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLysMetIle  103
      CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG

     GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCysG
 1948 GGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTG
```

FIG. 5B

lyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg 143
GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA

AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValP
2068 ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTAC

roValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 183
CAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
2188 AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAA

ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
AAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpG
2308 CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGG

luValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
AAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrA
2428 TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG

laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuG
2548 TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG

luProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
AGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
2668 TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATC

euLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
TGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluL
2788 GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAA

ysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 423
AAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
2908 AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCAC

euThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463 P
TAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA

IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluI
3028 TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAA

leGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503 O
TACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
3148 TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTAA

hrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543 L
CAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAAT

FIG. 5C

PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaT
3268 TTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCTA

hrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 583
CCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAGT

LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgG
3388 TAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGGG

luThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer 623
AGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCCA

IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspS
3508 TAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGATT

erGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 663
CGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCAC

GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLysG
3628 AACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAGG

luLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 703
AAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAG

AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAlaG
3748 ATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCCC

lnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu 743
AAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGC

ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluA
3868 CACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAG

laMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu 783
CCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTAG

GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluValI
3988 AAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTA

leProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp 823
TTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGGC

ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAlaA
4108 CAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCCG

laCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln 863
CCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAAG

GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGlnA
4228 GAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGG

laGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys 903
CTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAG

GlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleG
4348 GGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC

lnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg 943
AAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGG

FIG. 5D

```
     AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaV
4468 ACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCAG
     alValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle 983
     TAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATTA
     ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAspA
4588 GGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATT
     M
     AGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG
4708 TTTTATAGACATCACTATGAAAGTACTCATCCAAGAGTAAGTTCAGAAGTACACATCCCC
     CTAGGGGATGCTAAATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGAA
4828 TGGCATTTGGGCCAGGGAGTCGCCATAGAATGGAGGAAAAAGAAATATAGCACACAAGTA
     GACCCTGGCCTAGCAGACCAACTAATTCATCTGCATTATTTTGATTGTTTTTCAGAATCT
4948 GCTATAAAAAATGCCATATTAGGATATAGAGTTAGTCCTAGGTGTGAATATCAAGCAGGA
     CATAACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAAAAG
5068 ACAAAGCCACCTTTGCCTAGTGTTAAGAAACTGACAGAGGATAGATGGAACAAGCCCCAG
     AAGACCAAGGGCCACAGAGGGAGCCATACAATGAATGGACACTAGAGCTTTTAGAGGAGC
5188 TTAAGAGAGAAGCTGTTAGACATTTTCCTAGGCCATGGCTCCATAGCTTAGGACAATATA
     TCTATGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGCAAC
5308 AACTGCTGTTTATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTATTCAAC
     AGAGGAGAGCAAGAAGAAATGGAGCCAGTAGATCCTAATCTAGAGCCCTGGAAGCATCCA
5428 GGAAGTCAGCCTAGGACTGCTTGTAACAATTGCTATTGTAAAAAGTGTTGCTTTCATTGC
     TACGCGTGTTTCACAAGAAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAG
5548 CGACGAAGAGCTCCTCAGGACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAAGTA
     GTAAATGTAATGCAATCTTTACAAATATTAGCAATAGTATCATTAGTAGTAGTAGCAATA
5668 ATAGCAATAGTTGTGTGGACCATAGTACTCATAGAATATAGGAAAATATTAAGACAAAGA
                                                    MetLysVal    3
     AAATAGACAGATTAATTGATAGAATAAGAGAAAAAGCAGAAGACAGTGGCAATGAAAGTG
     LysGlyThrArgArgAsnTyrGlnHisLeuTrpArgTrpGlyThrLeuLeuLeuGlyMet
5788 AAGGGGACCAGGAGGAATTATCAGCACTTGTGGAGATGGGGCACCTTGCTCCTTGGGATG
     LeuMetIleCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal   43
     TTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG
     TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGlu
5908 TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAG
     ValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal   83
     GTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA
```

FIG. 5E

```
                ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMet
6028 GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATG

                GlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 123
     CAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA

                LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSerAsn
6148 CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAAT

                TrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 163
     TGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA

                ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAsp
6268 AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGAT

                AsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 203
     AATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC

                IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrPro
6388 ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCG

                AlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 243  E
     GCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA

                AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeu
6508 AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTG

                LeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 283  N
     TTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT

                AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsn
6628 GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAAC

                AsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 323  V
     AACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA

                IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThr
6748 ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACT

                LeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 363
     TTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT

                AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGlu
6868 AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAA

                PhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 403
     TTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA

                GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMet
6988 GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATG

                TrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 443
     TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA

                SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThr
7108 TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACC

                GluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 483
     GAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA
```

FIG. 5F

```
                TyrLysValIleLysIleGluProLeuGlyIleAlaProThrLysAlaLysArgArgVal
7228 TATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACCAAGGCAAAGAGAAGAGTG

                ValGlnArgGluLysArgAlaValGlyIleValGlyAlaMetPheLeuGlyPheLeuGly 523
     GTGCAGAGAGAAAAAAGAGCAGTGGGAATAGTAGGAGCTATGTTCCTTGGGTTCTTGGGA

                AlaAlaGlySerThrMetGlyAlaValSerLeuThrLeuThrValGlnAlaArgGlnLeu
7348 GCAGCAGGAAGCACTATGGGCGCAGTGTCATTGACGCTGACGGTACAGGCCAGACAATTA

                LeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHis 563
     TTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAACAT

                LeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGlu
7468 CTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTGGAA

                ArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIleCys 603
     AGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATTTGC

                ThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrpAsp
7588 ACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAAGACATTTGGGAT

                AsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyrThr 643
     AACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAACACAATATACACC

                LeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeuAsp
7708 TTACTTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTGGAT

                LysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpLeuTrpTyrIleLysIle 683
     AAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAAGATA

                PheIleMetIleValGlyGlyLeuValGlyLeuArgIleValPheAlaValLeuSerIle
7828 TTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTCTATA

                ValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrArgLeuProValPro 723
     GTGAATAGAGTTAGGCAGGGATACTCACCATTGTCATTTCAGACCCGCCTCCCAGTCCCG

                ArgGlyProAspArgProAspGlyIleGluGluGluGlyGlyGluArgAspArgAspArg
7948 AGGGGACCCGACAGGCCCGACGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA

                SerValArgLeuValAspGlyPheLeuAlaLeuIleTrpGluAspLeuArgSerLeuCys 763
     TCCGTTCGATTAGTGGATGGATTCTTAGCACTTATCTGGGAAGATCTGCGGAGCCTGTGC

                LeuPheSerTyrArgArgLeuArgAspLeuLeuLeuIleAlaAlaArgThrValGluIle
8068 CTCTTCAGCTACCGCCGCTTGAGAGACTTACTCTTGATTGCAGCGAGGACTGTGGAAATT

                LeuGlyHisArgGlyTrpGluAlaLeuLysTyrTrpTrpSerLeuLeuGlnTyrTrpIle 803
     CTGGGGCACAGGGGGTGGGAAGCCCTCAAATATTGGTGGAGTCTCCTGCAGTATTGGATT

                GlnGluLeuLysAsnSerAlaValSerTrpLeuAsnAlaThrAlaIleAlaValThrGlu
8188 CAGGAACTAAAGAATAGTGCTGTTAGCTGGCTCAACGCCACAGCTATAGCAGTAACTGAG

                GlyThrAspArgValIleGluValAlaGlnArgAlaTyrArgAlaIleLeuHisIleHis 843
     GGGACAGATAGGGTTATAGAAGTAGCACAAAGAGCTTATAGAGCTATTCTCCACATACAT

                ArgArgIleArgGlnGlyLeuGluArgLeuLeuLeuOC
8308 AGAAGAATTAGACAGGGCTTGGAAAGGCTTTTGCTATAAGATGGGTGGCAAGTGGTCAAA

     ACGTAGTATGGGTGGATGGTCTGCTATAAGGGAAAGAATGAGACGAGCTGAGCCACGAGC
```

FIG. 5G

```
8428 TGAGCCAGCAGCAGATGGGGTGGGAGCAGTATCTCGAGACCTGGAAAAACATGGAGCAAT
     CACAAGTAGCAATACAGCAGCTACTAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGGA
8548 GGAAGAGGTGGGTTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTACAAGGC
                                                U3→
     AGCTTTAGATATTAGCCACTTTTTAAAAGAAAAGGGGGGA CTGGAAGGGCTAATTTGGT   L
8667 CCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTG
     ATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCT   T
8787 TCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAACA
     GCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGAGAAAGAAGTGTTAGTGT
8907 GGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACT   R
     ACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGA
9027 GGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGCTGCATATAAGCAGCTGC
          ←U3 R→
     TTTTTGCCTGTACTG GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGC
                                                ←R U5→
9146 TAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCA AGTAGT
     GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGT
          ←U5
9265 GTGGAAAAATCTCTAGCAG
```

FIG. 5H

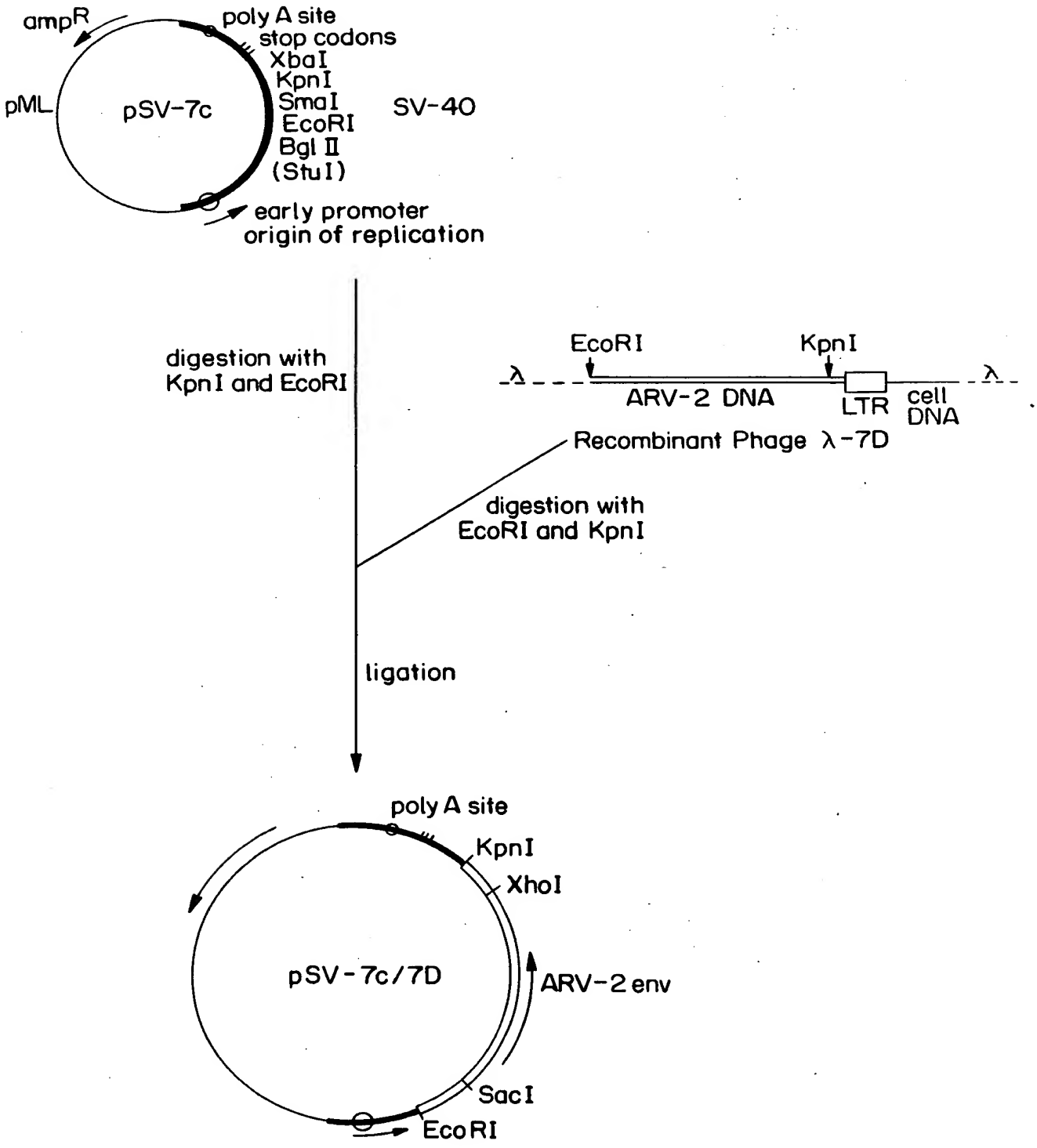
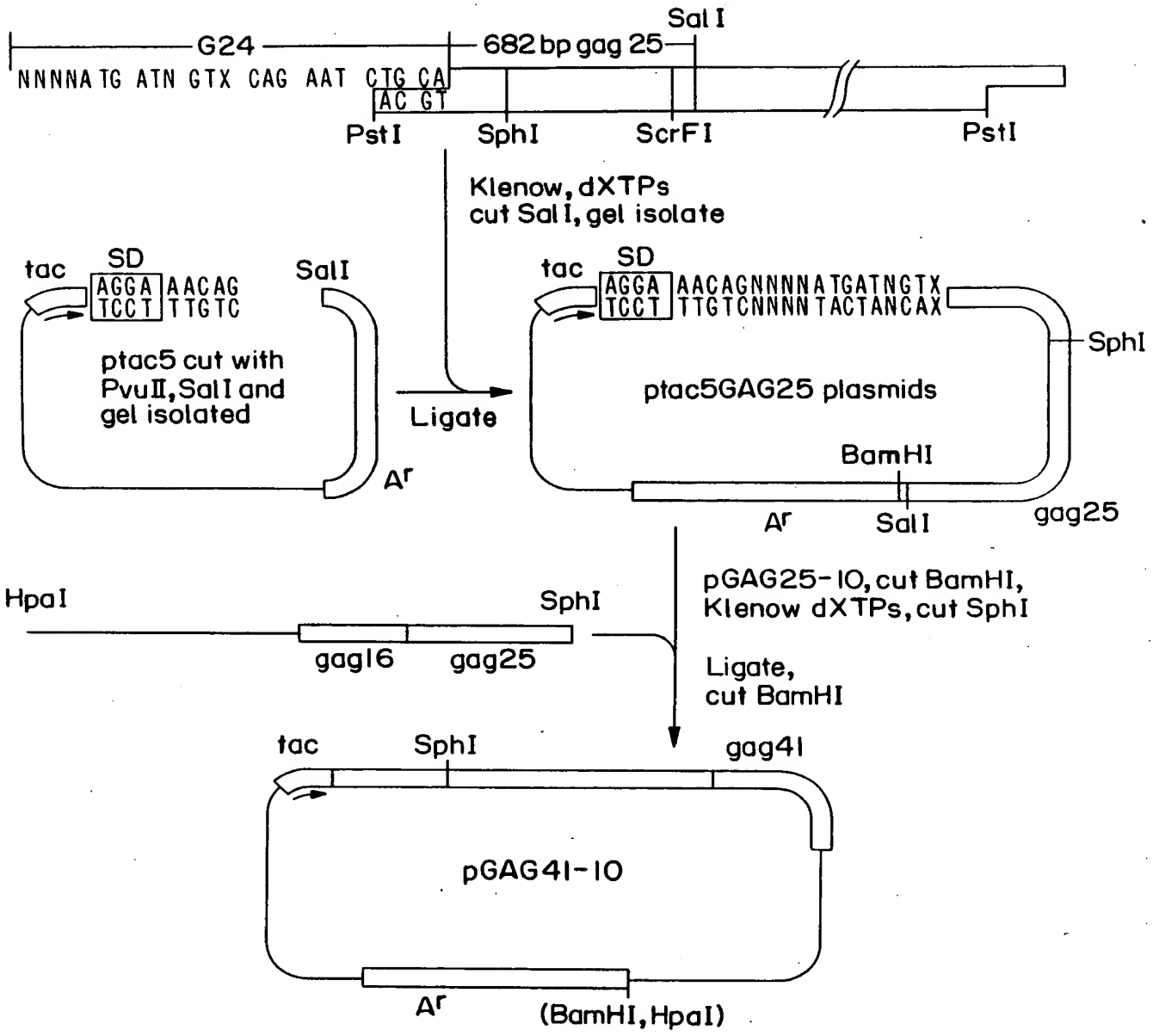

ampR
poly A site
stop codons
XbaI
KpnI
SmaI
EcoRI
Bgl II
(StuI)
pML
pSV-7c
SV-40
early promoter
origin of replication
digestion with
KpnI and EcoRI
EcoRI
KpnI
λ
ARV-2 DNA
LTR
cell
DNA
λ
Recombinant Phage λ-7D
digestion with
EcoRI and KpnI
ligation
poly A site
KpnI
XhoI
pSV-7c/7D
ARV-2 env
SacI
EcoRI

FIG. 6

FIG. 7


G24
682 bp gag 25
Sal I
NNNNA TG ATN GTX CAG AAT CTG CA
AC GT
PstI
SphI
ScrFI
PstI
Klenow, dXTPs
cut Sal I, gel isolate
tac
SD
AGGA
TCCT
AACAG
TTGTC
SalI
ptac5 cut with
PvuII, SalI and
gel isolated
Ar
Ligate
tac
SD
AGGA
TCCT
AACAGNNNNATGATNGTX
TTGTCNNNNTACTANCAX
SphI
ptac5GAG25 plasmids
BamHI
Ar
SalI
gag25
pGAG25-10, cut BamHI,
Klenow dXTPs, cut SphI
HpaI
SphI
gag16
gag25
Ligate,
cut BamHI
tac
SphI
gag41
pGAG41-10
Ar
(BamHI, HpaI)

01 931154

```
                                                            *
                       ptac 5 Promotor                    MetIleVal
                                                          ATGATCGTA

     GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
748  CAGAATCTGCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG

     ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu  181
     GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA

     SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
868  TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA

     AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal  221
     GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG

     HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
988  CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC

     IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro  261
     ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT

     IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA

     MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp  301
     ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT

     TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT

     TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys  341
     TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA

     AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA

     ProGlyHisLysAlaArgValLeu  Stop  Stop
     CCCGGGCATAAAGCAAGAGTTTTGTGATAG          ptac 5
```

FIG. 8

ptac 5 Promotor MetIleVal 141
ATGATCGTA

GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
748 CAGAATCTGCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG

ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu 181
GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA G

SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA

AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal 221
GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG

HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp A
988 CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC

IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro 261
ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT

IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA G

MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp 301
ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT

TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT

TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys 341
TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA

AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA

ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn 381
CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC

IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1468 ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT

GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg 421
GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA

CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
PhePheArgG
1588 TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG

LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro 461
luAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAla 23
AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA

ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
AsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSerLeuSerGluA
1708 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG

GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn 501
laGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGln 63
CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC

FIG. 9A

```
      AspProSerSerGlnOC
     ArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyA
1828 GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAG

      laAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLysMetIle 103
      CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG

     GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCysG
1948 GGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTG

      lyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg 143
      GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA

     AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValP
2068 ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTAC

      roValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 183
      CAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

     GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
2188 AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAA

      ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
      AAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

     ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpG
2308 CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGG

      luValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
      AAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

     LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrA
2428 TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG

      laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
      CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

     LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuG
2548 TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG

      luProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
      AGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

     ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
2668 TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATC

      euLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
      TGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

     TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluL
2788 GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAA

      ysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 423
      AAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

     GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
2908 AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCAC

      euThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463
      TAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA
```

POL

FIG. 9B

IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluI
3028 TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAA

leGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503
TACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
3148 TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTT

hrGluAlaValGluLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543
ptac 5

FIG. 9C

ARV GAG p16 - synthetic Parts A and B

```
5'      arv 234          3'
     MetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLys
TATTATGCAAAGAGGTAACTTCAGGAATCAAAGAAAGACCGTTAAGTGTTTCAACTGTGGTAAG
ATAATACGTTTCTCCATTGAAGTCCTTAGTTTCTTTCTGGCAATTCACAAAGTTGACACCATTC
3'             ^ arv 235    ^ 5'
     10 mnll, 23 hinfl,

     GluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysAlaCysTrpArgCysGly
 63  GAAGGTCACATCGCTAAGAACTGTAGAGCTCCAAGAAAGAAGGCTTGTTGGAGATGTGGT
     CTTCCAGTGTAGCGATTCTTGACATCTCGAGGTTCTTTCTTCCGAACAACCTCTACACCA

     76 ddel, 88 ban2 hglA hglJ11 sacl sduI, 89 alul,

     ArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIle
123  AGAGAAGGTCACCAAATGAAGGACTGTACCGAAAGACAAGCTAACTTCTTGGGTAAGATC
     TCTCTTCCAGTGGTTTACTTCCTGACATGGCTTTCTGTTCGATTGAAGAACCCATTCTAG

     129 bstE2, 131 hph, 148 rsaI, 161 alul, 178 bglll xho2, 179
     sau3a,

     TrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAla
183  TGGCCATCTTACAAGGGTAGACCAGGTAACTTCTTGCAATCCAGACCAGAACCAACCGCT
     ACCGGTAGAATGTTCCCATCTGGTCCATTGAAGAACGTTAGGTCTGGTCTTGGTTGGCGA

     183 ball cfrl hael, 184 haelll, 199 accl, 204 apyl ecorll sc
     rFl,

     ProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGlu
243  CCACCTGAAGAAAGTTTCAGGTTCGGTGAAGAAAAGACCACCCCATCTCAAAAGCAAGAA
     GGTGGACTTCTTTCAAAGTCCAAGCCACTTCTTTTCTGGTGGGGTAGAGTTTTCGTTCTT

      249 mboll, 267 hph, 270 mboll,

     ProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspPro
303  CCAATCGACAAGGAATTGTACCCATTGACCTCTTTGAGATCCTTGTTCGGTAACGATCCC
     GGTTAGCTGTTCCTTAACATGGGTAACTGGAGAAACTCTAGGAACAAGCCATTGCTAGGG

     307 taql, 320 rsaI, 331 mnll, 339 xho2, 340 sau3a, 357 sau3a,
     361 mnll, 362 aval xhol,

     SerSerGlnOP AM
363  TCGAGCCAATGATAG
     AGCTCGGTTACTATCAGCT

     363 taql, 377 accl hindll sall
```

FIG. 10

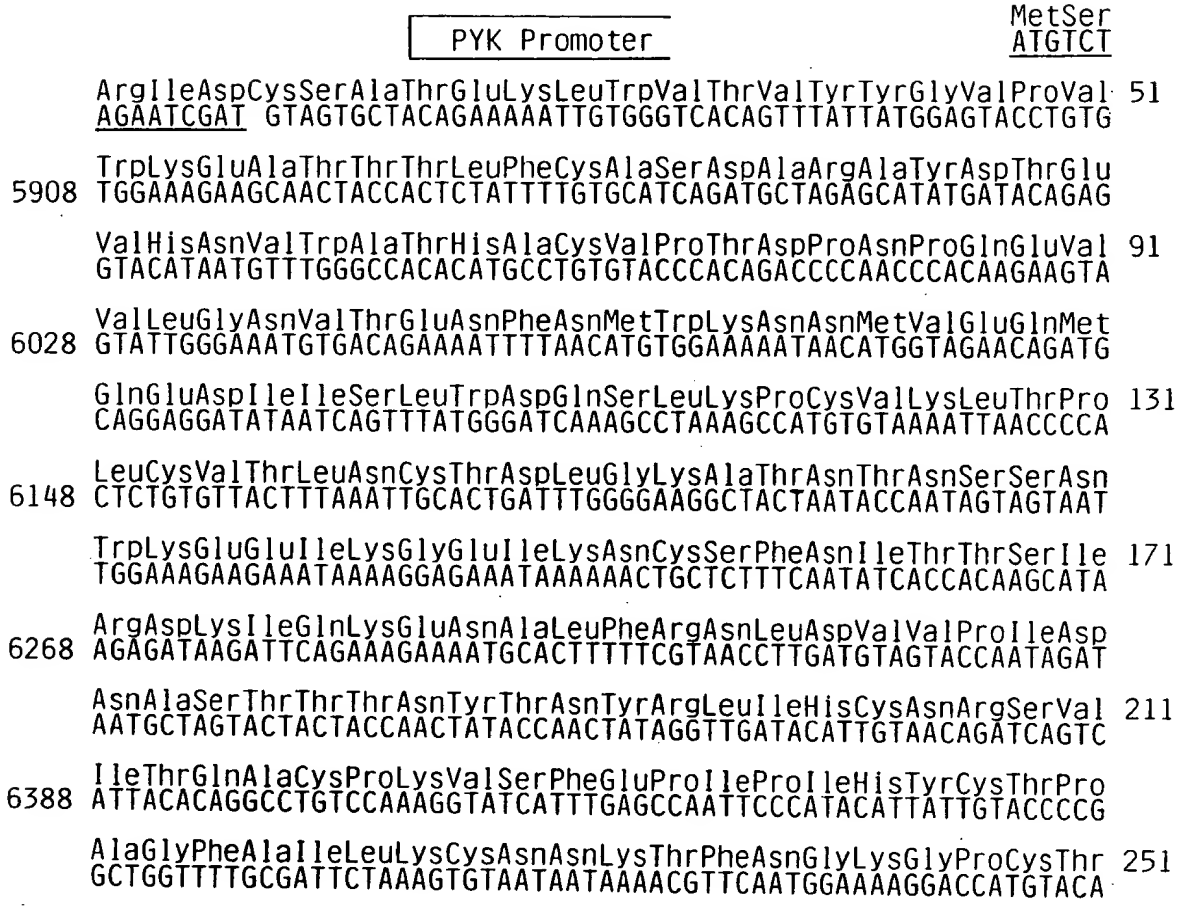

PYK Promoter
MetSer
ATGTCT
ArgIleAspCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal 51
AGAATCGAT GTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG
TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGlu
5908 TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAG
ValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal 91
GTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA
ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMet
6028 GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATG
GlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 131
CAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA
LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSerAsn
6148 CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAAT
TrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 171
TGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA
ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAsp
6268 AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGAT
AsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 211
AATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC
IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrPro
6388 ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCG
AlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 251
GCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA

FIG. 11A

```
                AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeu
6508 AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTG

                LeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 291
     TTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT

                AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsn
6628 GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAAC

                AsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 331
     AACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA

                IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThr
6748 ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACT

                LeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 371
     TTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT

                AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGlu
6868 AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAA

                PhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 411
     TTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA

                GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMet
6988 GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATG

                TrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 451
     TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA

                SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThr
7108 TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACC

                GluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 491
     GAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA

                TyrLysValIleLysIleGluProAsnSerValSer
7228 TATAAAGTAATAAAAATTGAACCAAATTCGGTATCTTGA
```

PYK Terminator

E

N

V

FIG. 11B

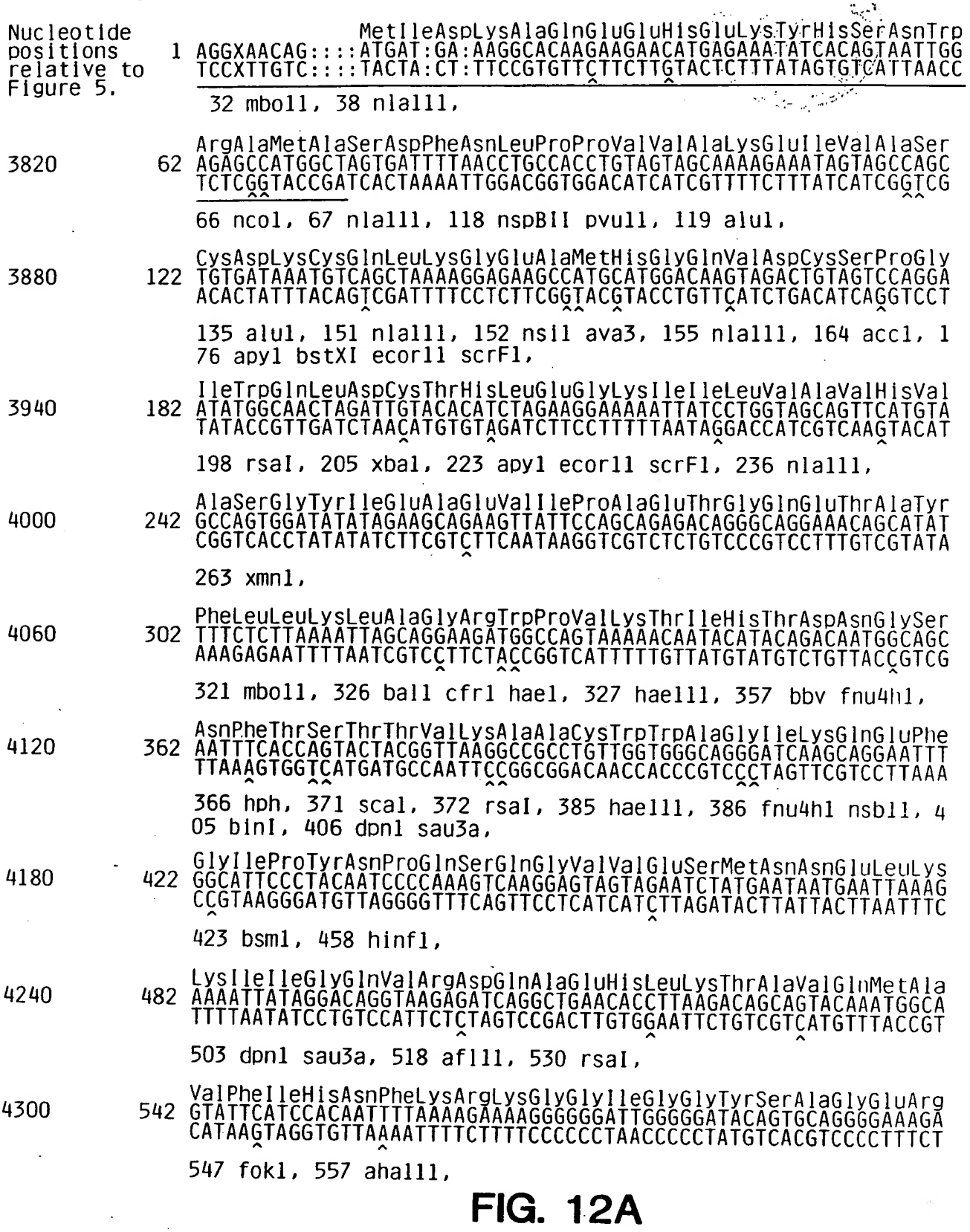
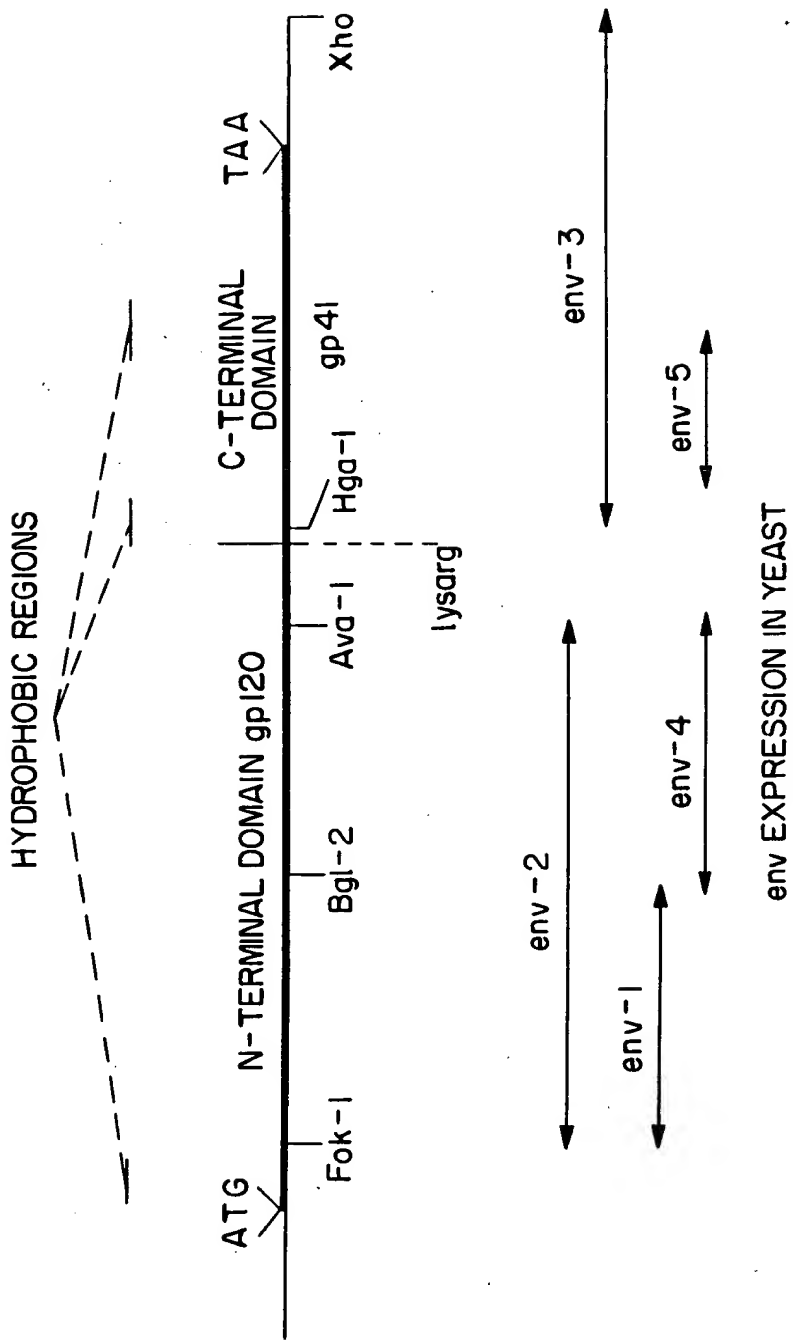

Nucleotide positions relative to Figure 5.
1
MetIleAspLysAlaGlnGluGluHisGluLysTyrHisSerAsnTrp
AGGXAACAG::::ATGAT:GA:AAGGCACAAGAAGAACATGAGAAATATCACAGTAATTGG
TCCXTTGTC::::TACTA:CT:TTCCGTGTTCTTCTTGTACTCTTTATAGTGTCATTAACC
32 mboll, 38 nlalll,
3820
62
ArgAlaMetAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSer
AGAGCCATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC
TCTCGGTACCGATCACTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCG
66 ncol, 67 nlalll, 118 nspBII pvull, 119 alul,
3880
122
CysAspLysCysGlnLeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGly
TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA
ACACTATTTACAGTCGATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCT
135 alul, 151 nlalll, 152 nsil ava3, 155 nlalll, 164 acc1, 1
76 apyl bstXI ecorll scrFl,
3940
182
IleTrpGlnLeuAspCysThrHisLeuGluGlyLysIleIleLeuValAlaValHisVal
ATATGGCAACTAGATTGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTA
TATACCGTTGATCTAACATGTGTAGATCTTCCTTTTTAATAGGACCATCGTCAAGTACAT
198 rsaI, 205 xbal, 223 apyl ecorll scrFl, 236 nlalll,
4000
242
AlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyr
GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATAT
CGGTCACCTATATATCTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATA
263 xmnl,
4060
302
PheLeuLeuLysLeuAlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySer
TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGC
AAAGAGAATTTTAATCGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCG
321 mboll, 326 ball cfrl hael, 327 haelll, 357 bbv fnu4hl,
4120
362
AsnPheThrSerThrThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPhe
AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTT
TTAAAGTGGTCATGATGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAA
366 hph, 371 scal, 372 rsaI, 385 haelll, 386 fnu4hl nsbll, 4
05 binI, 406 dpnl sau3a,
4180
422
GlyIleProTyrAsnProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLys
GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAG
CCGTAAGGGATGTTAGGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTC
423 bsml, 458 hinfl,
4240
482
LysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAla
AAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCA
TTTTAATATCCTGTCCATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGT
503 dpnl sau3a, 518 afllI, 530 rsaI,
4300
542
ValPheIleHisAsnPheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArg
GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGA
CATAAGTAGGTGTTAAAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCT
547 fokl, 557 ahalll,

FIG. 12A

```
                IleValAspIleIleAlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLys
4360        602 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAAGCAAATTACAAAA
                TATCATCTGTATTATCGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTT
                    ^
                605 acc1,

                IleGlnAsnPheArgValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAla
4420        662 ATTCAAAATTTTCGGGTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCA
                TAAGTTTTAAAAGCCCAAATAATGTCCCTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGT
                                              ^^          ^
                697 xho2, 698 dpn1 sau3a, 713 asu1 ava2,

                LysLeuLeuTrpLysGlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysVal
4480        722 AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA
                TTCGAAGAGACCTTTCCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCAT
                ^^          ^
                722 hind111, 723 alu1, 737 hph,

                ValProArgArgLysAlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAsp
4540        782 GTGCCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGAT
                CACGGTTCTTCTTTTCGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTA
                    ^                                           ^
                789 mbo11, 833 hph,

                CysValAlaSerArgGlnAspGluAspAM
4600        842 TGTGTGGCAAGTAGACAGGATGAGGATTAGTCGACGGAATTCTTTAGTAAAACACC
                ACACACCGTTCATCTGTCCTACTCCTAATCAGCTGCCTTAAGAAATCATTTTGTGG
                    ^       ^   ^   ^   ^
                852 acc1, 859 fok1, 863 mn11, 871 acc1 hind11 sal1, 872 taq1
                , 878 ecor1,
```

FIG. 12B

FIG. 13


HYDROPHOBIC REGIONS
ATG
Fok-1
N-TERMINAL DOMAIN gp120
Bgl-2
Ava-1
lysarg
Hga-1
C-TERMINAL DOMAIN
gp41
TAA
Xho
env-1
env-2
env-4
env-5
env-3
env EXPRESSION IN YEAST AND BACTERIA

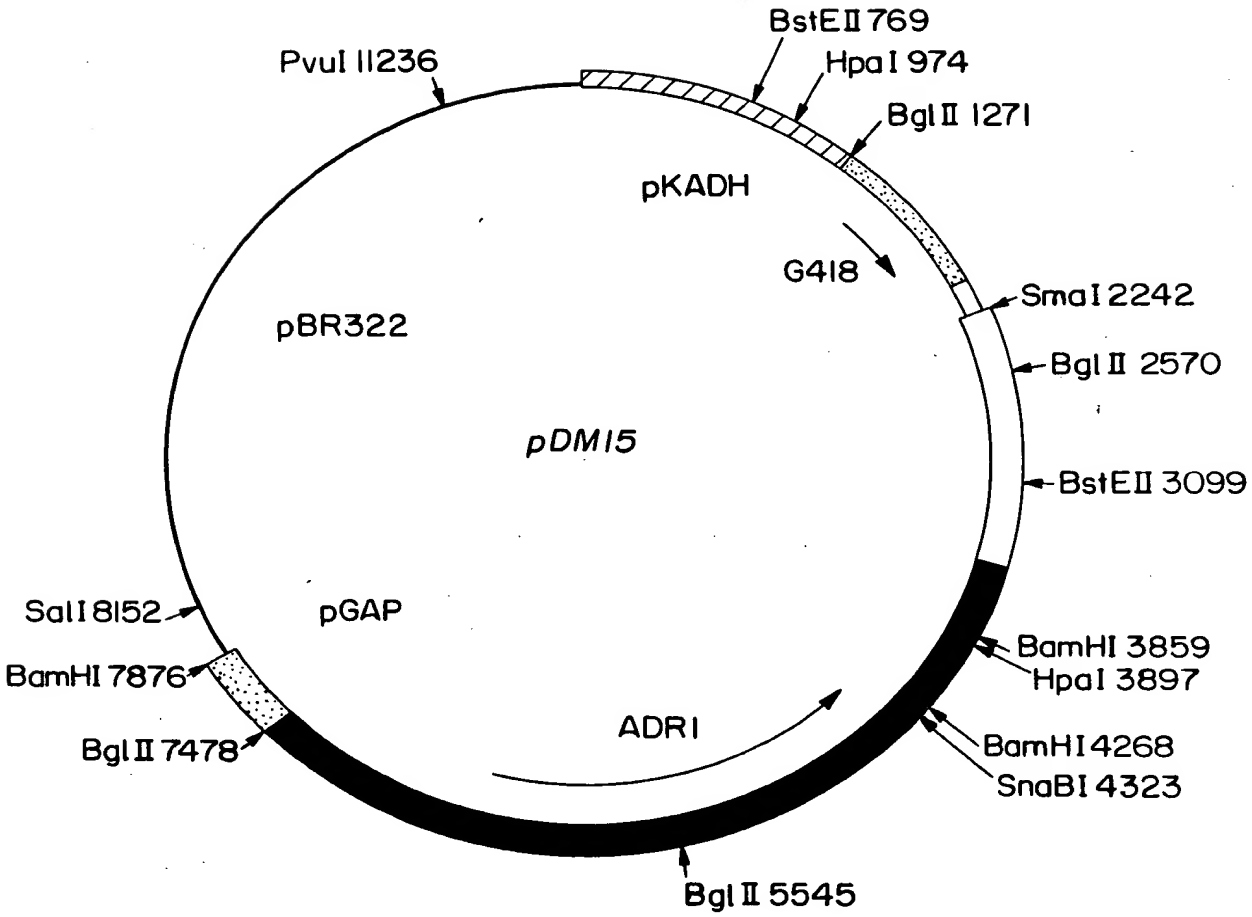

BstEII 769
HpaI 974
BglII 1271
PvuI 11236
pKADH
G418
SmaI 2242
pBR322
BglII 2570
pDM15
BstEII 3099
SalI 8152
pGAP
BamHI 3859
HpaI 3897
BamHI 7876
ADR1
BamHI 4268
SnaBI 4323
BglII 7478
BglII 5545

FIG. 14

```
                SOD
               MetAlaThrLysAlaValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsn
   1 CATGGCGACGAAGGCCGTGTGCGTGCTGAAGGGCGACGGCCCAGTGCAGGGCATCATCAAT
       CGCTGCTTCCGGCACACGCACGACTTCCCGCTGCCGGGTCACGTCCCGTAGTAGTTA

               PheGluGlnLysGluSerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThr
  62 TTCGAGCAGAAGGAAAGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACT
     AAGCTCGTCTTCCTTTCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGA

               GluGlyLeuHisGlyPheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSer
 122 GAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGT
     CTTCCGGACGTACCTAAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCA

               AlaGlyProHisPheAsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArg
 182 GCAGGTCCTCACTTTAATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGG
     CGTCCAGGAGTGAAATTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCC

               HisValGlyAspLeuGlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIle
 242 CATGTTGGAGACTTGGGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATT
     GTACAACCTCTGAACCCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAA

               GluAspSerValIleSerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValVal
 302 GAAGATTCTGTGATCTCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTC
     CTTCTAAGACACTAGAGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAG

               HisGluLysAlaAspAspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsn
 362 CATGAAAAAGCAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAAC
     GTACTTTTTCGTCTACTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTG

                                                        ENV 5B
               AlaGlySerArgLeuAlaCysGlyValIleGlyIleAlaMetAlaIleGluAlaGlnGln
 422 GCTGGAAGTCGTTTGGCTTGTGGTGTAATTGGGATCGCCATGGCTATCGAAGCTCAACAA
     CGACCTTCAGCAAACCGAACACCACATTAACCCTAGCGGTACCGATAGCTTCGAGTTGTT

               HisLeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaVal
 482 CACTTGCTGCAGTTGACCGTTTGGGGTATCAAGCAGTTGCAGGCTAGAGTTTTGGCTGTT
     GTGAACGACGTCAACTGGCAAACCCCATAGTTCGTCAACGTCCGATCTCAAAACCGACAA

               GluArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIle
 542 GAAAGATACTTGAGAGATCAACAATTGTTGGGTATCTGGGGTTGTTCTGGTAAGTTGATT
     CTTTCTATGAACTCTCTAGTTGTTAACAACCCATAGACCCCAACAAGACCATTCAACTAA

               CysThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrp
 602 TGTACCACCGCTGTTCCCTGGAACGCTTCTTGGTCTAACAAGTCTTTGGAAGACATCTGG
     ACATGGTGGCGACAAGGGACCTTGCGAAGAACCAGATTGTTCAGAAACCTTCTGTAGACC

               AspAsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyr
 662 GACAACATGACCTGGATGCAATGGGAAAGAGAAATCGACAACTACACCAACACCATCTAC
     CTGTTGTACTGGACCTACGTTACCCTTTCTCTTTAGCTGTTGATGTGGTTGTGGTAGATG

               ThrLeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeu
 722 ACCTTGTTGGAGGAATCTCAAAACCAACAAGAAAAGAACGAACAAGAATTGTTGGAATTG
     TGGAACAACCTCCTTAGAGTTTTGGTTGTTCTTTTCTTGCTTGTTCTTAACAACCTTAAC

               AspLysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpAM
 782 GACAAGTGGGCAAGCTTGTGGAACTGGTTCTCTATCACCAACTGGTAG
     CTGTTCACCCGTTCGAACACCTTGACCAAGAGATAGTGGTTGACCATCAGCT

          Translated Mol. Weight = 30414.22
```

FIG. 15

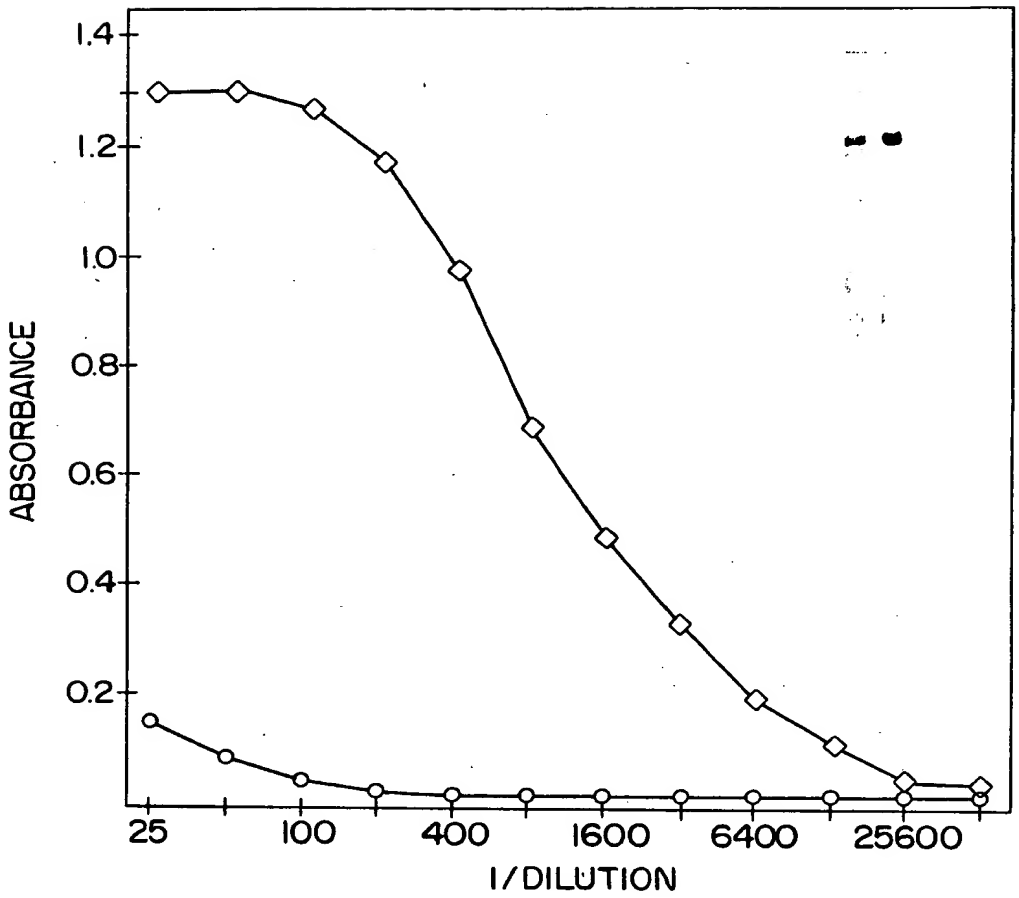

1.4
1.2
1.0
0.8
0.6
0.4
0.2
ABSORBANCE
25
100
400
1600
6400
25600
1/DILUTION

FIG. 16A

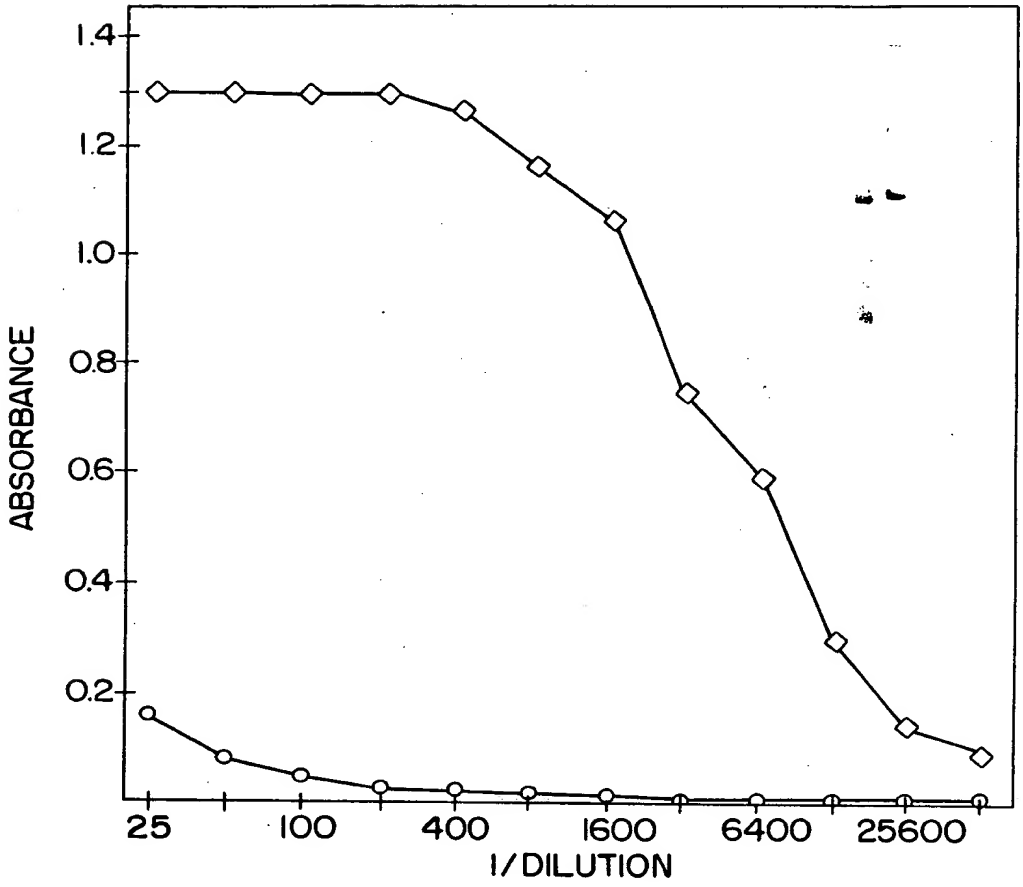

1.4
1.2
1.0
0.8
0.6
0.4
0.2
ABSORBANCE
25
100
400
1600
6400
25600
1/DILUTION

FIG. 16B

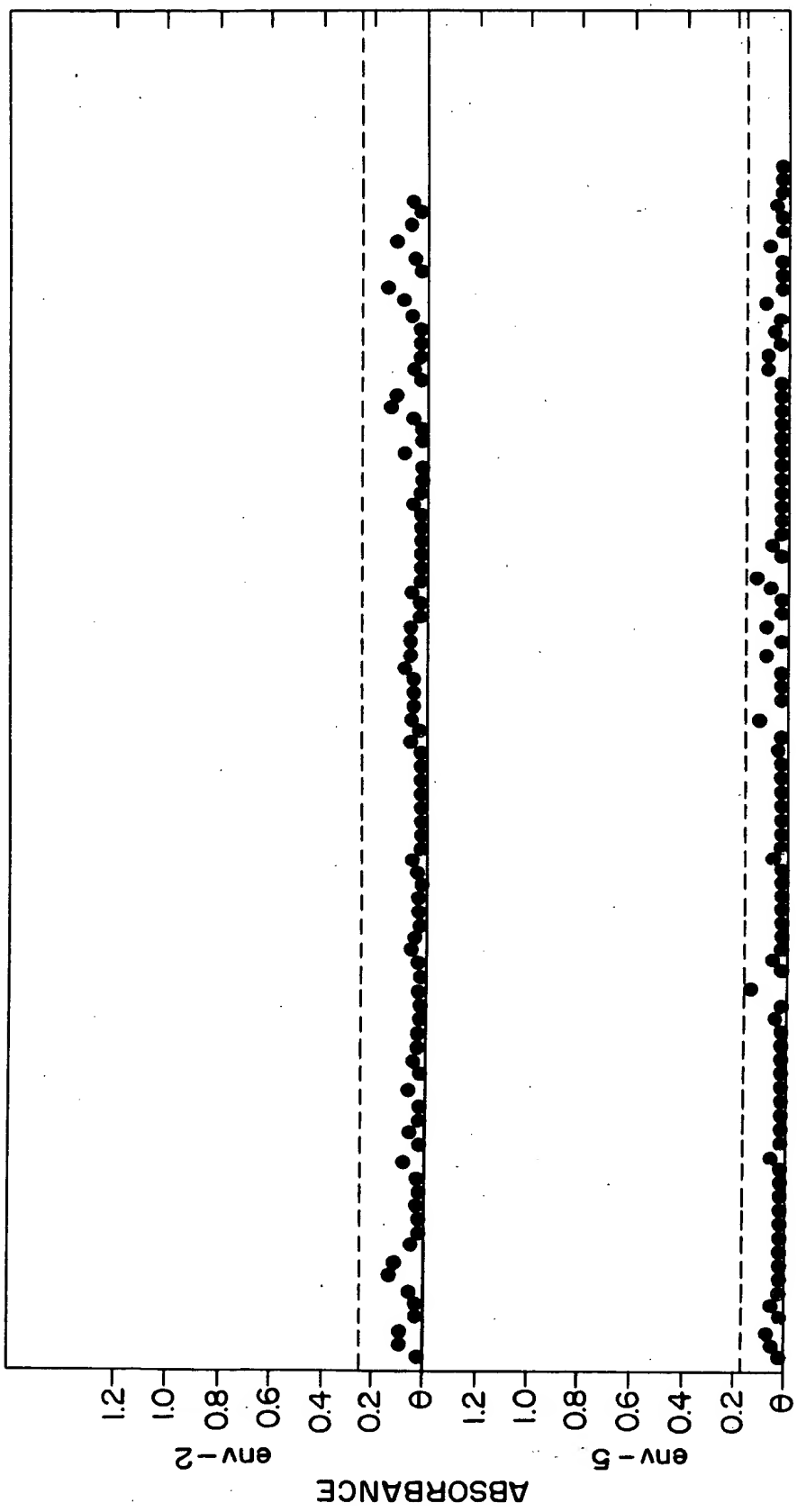


FIG. 17

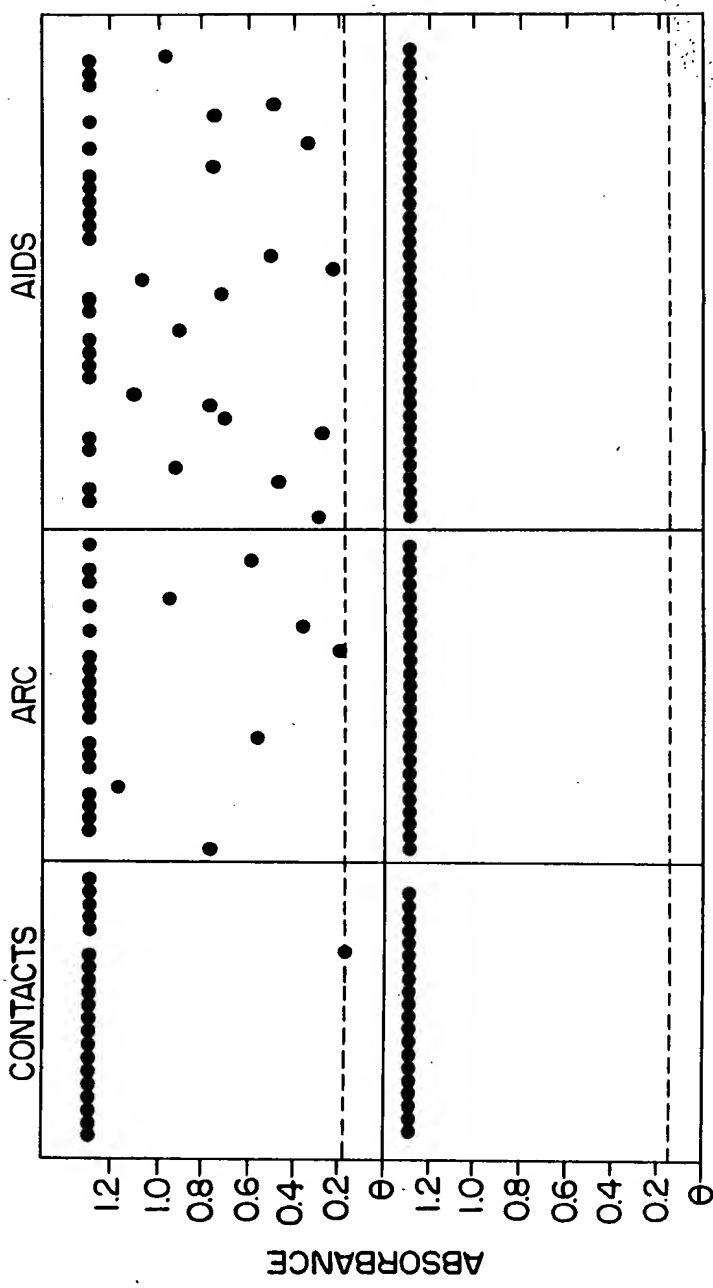


FIG. 18

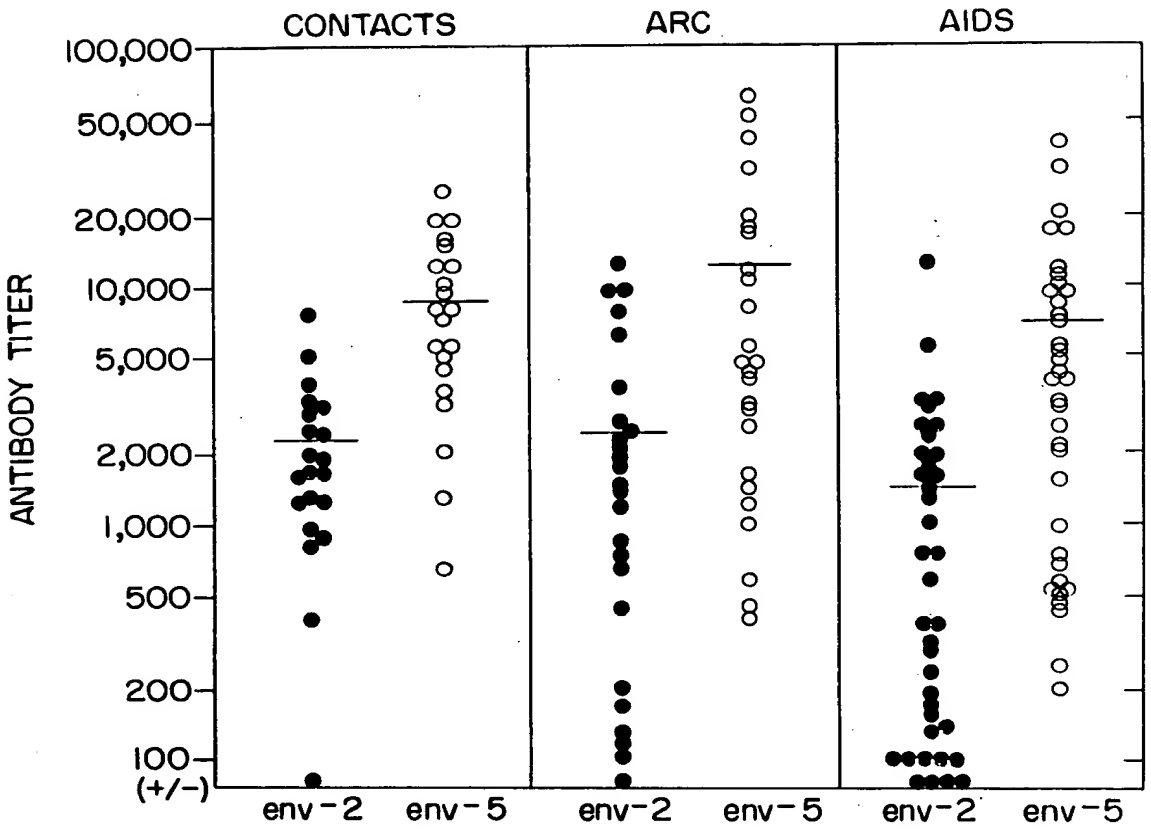

CONTACTS
ARC
AIDS
ANTIBODY TITER
100,000
50,000
20,000
10,000
5,000
2,000
1,000
500
200
100
(+/-)
env-2
env-5
env-2
env-5
env-2
env-5

FIG. 19

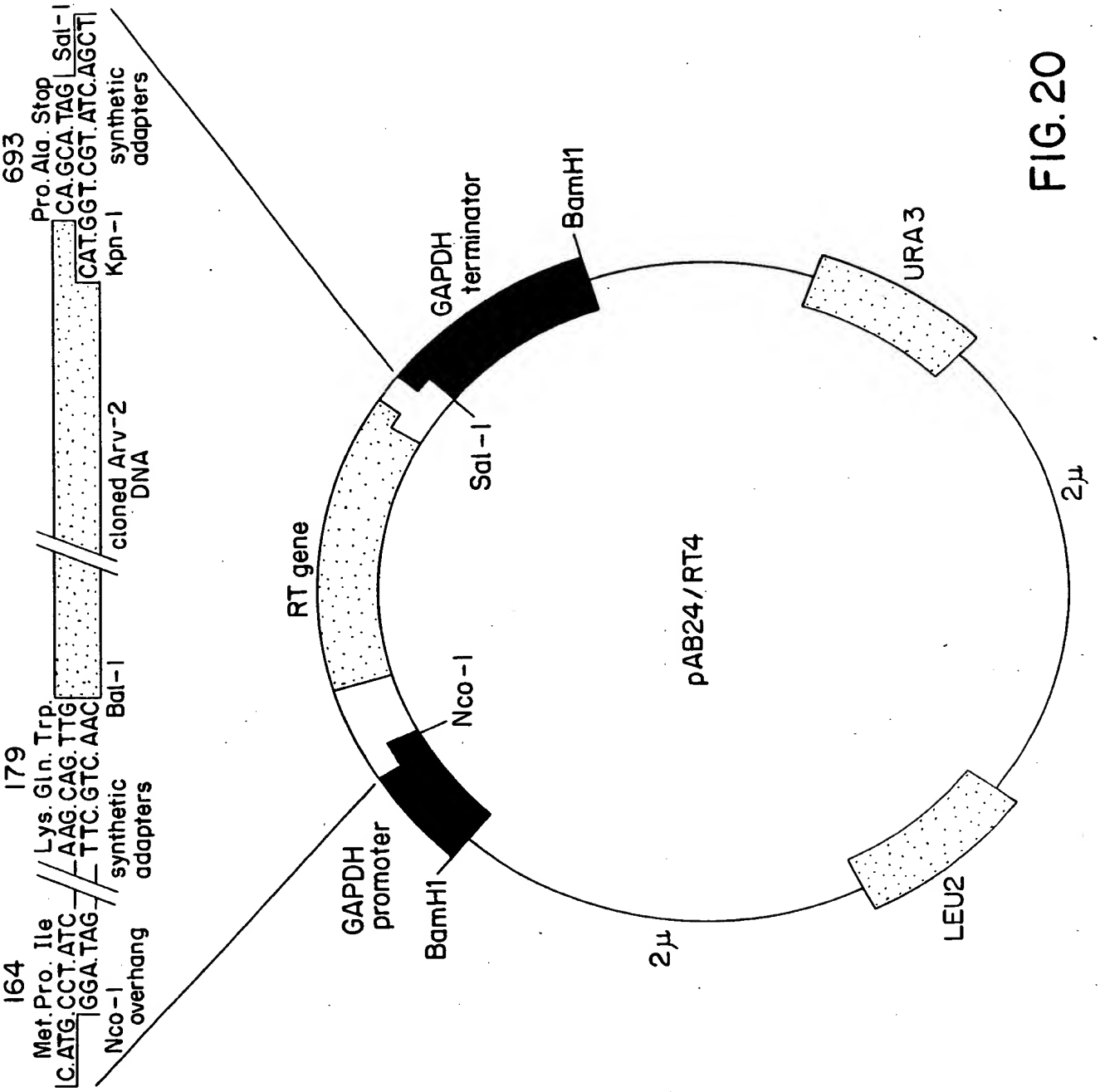

693
Pro. Ala. Stop
CA.GCA.TAG
Sal-I
CAT.GGT.CGT.ATC.AGCT
Kpn-I
synthetic adapters
cloned Arv-2 DNA
179
Lys. Gln. Trp.
AAG.CAG.TTG
TTC.GTC.AAC
Bal-I
synthetic adapters
164
Met. Pro. Ile
C.ATG.CCT.ATC
GGA.TAG
Nco-I overhang
GAPDH promoter
BamH1
2μ
LEU2
Nco-I
RT gene
pAB24/RT4
Sal-I
GAPDH terminator
BamH1
URA3
2μ

FIG. 20

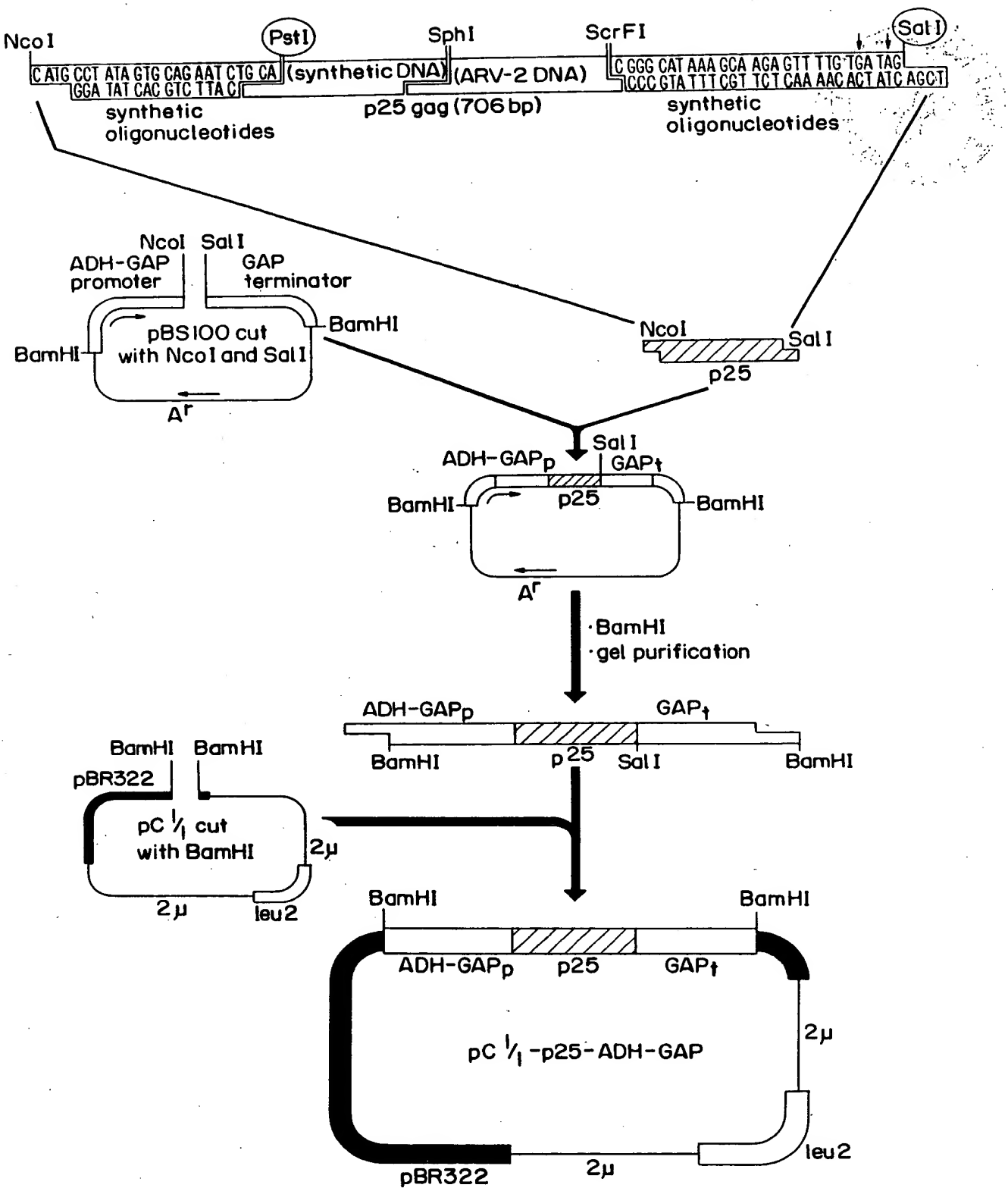
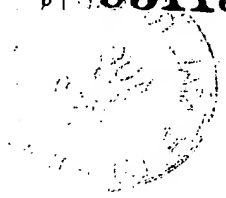

NcoI
PstI
SphI
ScrFI
SalI
C ATG CCT ATA GTG CAG AAT CTG CA
GGA TAT CAC GTC TTA C
(synthetic DNA)
(ARV-2 DNA)
C GGG CAT AAA GCA AGA GTT TTG TGA TAG
CCC GTA TTT CGT TCT CAA AAC ACT ATC AGCT
synthetic oligonucleotides
p25 gag (706 bp)
synthetic oligonucleotides
NcoI
SalI
ADH-GAP promoter
GAP terminator
BamHI
pBS100 cut with NcoI and SalI
BamHI
Ar
NcoI
SalI
p25
SalI
ADH-GAPp
GAPt
BamHI
p25
BamHI
Ar
·BamHI
·gel purification
ADH-GAPp
GAPt
BamHI
p25
SalI
BamHI
BamHI
BamHI
pBR322
pC 1/1 cut with BamHI
2μ
2μ
leu2
BamHI
BamHI
ADH-GAPp
p25
GAPt
pC 1/1 -p25-ADH-GAP
2μ
leu2
pBR322
2μ

FIG. 21

```
     1                                                  10
     Met Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln
C    ATG CCT ATA GTG CAG AAT CTG CAG GGG CAA ATG GTA CAT CAG

                    20
Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
GCC ATA TCA CCT AGA ACT TTA AAT GCT TGG GTA AAA GTA GTA GAA

 30                                                 40
Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu
GAA AAG GCT TTC AGC CCA GAA GTA ATA CCC ATG TTT TCA GCA TTA

                    50
Ser Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr
TCA GAA GGA GCC ACC CCT CAA GAT TTA AAC ACC ATG CTA AAC ACA

 60                                                 70
Val Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile
GTG GGG GGA CAT CAA GCA GCC ATG CAA ATG TTA AAA GAG ACT ATC

                    80
Asn Glu Glu Ala Ala Glu Trp Asp Arg Val His Pro Val His Ala
AAT GAG GAG GCT GCC GAA TGG GAT AGA GTG CAT CCA GTG CAT GCA

 90                                                100
Gly Pro Ile Ala Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp
GGG CCT ATT GCA CCA GGC CAA ATG AGA GAA CCA AGG GGA AGT GAC
```

FIG. 22A



```
                                110
Ile Ala Gly Thr Thr Ser Thr Leu Gln Glu Gln Ile Gly Trp Met
ATA GCA GGA ACT ACT AGT ACC CTT CAG GAA CAA ATA GGA TGG ATG

120                                                 130
Thr Asn Asn Pro Pro Ile Pro Val Gly Glu Ile Tyr Lys Arg Trp
ACA AAT AAT CCA CCT ATC CCA GTA GGA GAA ATC TAT AAA AGA TGG

                    140
Ile Ile Leu Gly Leu Asn Lys Ile Val Arg Met Tyr Ser Pro Thr
ATA ATC CTG GGA TTA AAT AAA ATA GTA AGA ATG TAT AGC CCT ACC

150                                                 160
Ser Ile Leu Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp
AGC ATT CTG GAC ATA AGA CAA GGA CCA AAG GAA CCC TTT AGA GAT

                    170
Tyr Val Asp Arg Phe Tyr Lys Thr Leu Arg Ala Glu Gln Ala Ser
TAT GTA GAC CGG TTC TAT AAA ACT CTA AGA GCC GAA CAA GCT TCA

180                                                 190
Gln Asp Val Lys Asn Trp Met Thr Glu Thr Leu Leu Val Gln Asn
CAG GAT GTA AAA AAT TGG ATG ACA GAA ACC TTG TTG GTC CAA AAT

                    200
Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala Leu Gly Pro Ala
GCA AAC CCA GAT TGT AAG ACT ATT TTA AAA GCA TTG GGA CCA GCA

210                                                 220
Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
GCT ACA CTA GAA GAA ATG ATG ACA GCA TGT CAG GGA GTG GGG GGA

                    230     232
Pro Gly His Lys Ala Arg Val Leu OP
CCC GGG CAT AAA GCA AGA GTT TTG TGA TAG
```

Translated Mol. Weight = 25700.75

FIG. 22B

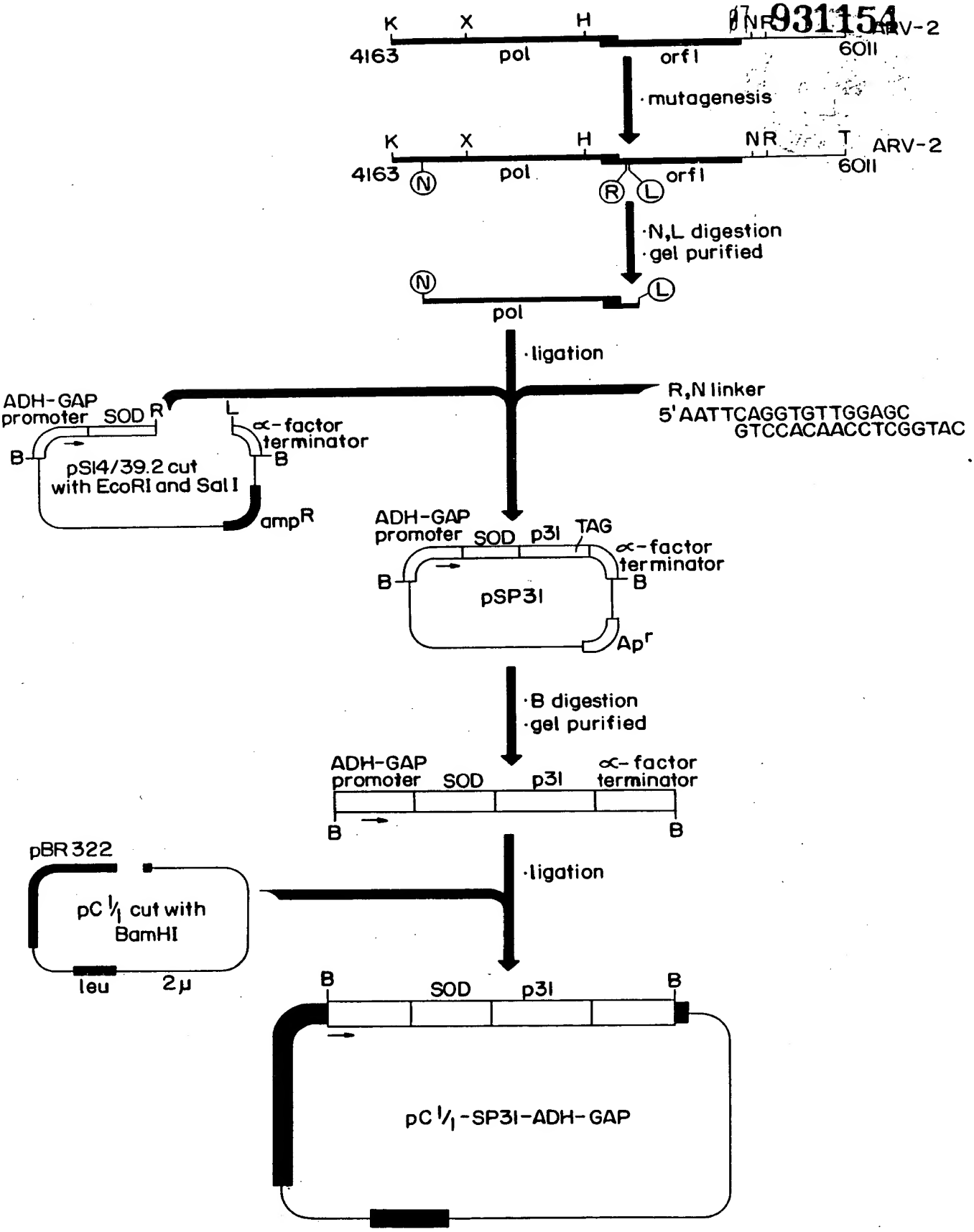

K
X
H
NR
T
ARV-2
4163
pol
orf1
6011
·mutagenesis
N
R
L
·N,L digestion
·gel purified
·ligation
R,N linker
5'AATTCAGGTGTTGGAGC
GTCCACAACCTCGGTAC
ADH-GAP promoter
SOD
R
L
α-factor terminator
B
pSI4/39.2 cut with EcoRI and SalI
ampR
p31
TAG
pSP31
Apr
·B digestion
·gel purified
pBR322
pC 1/1 cut with BamHI
leu
2μ
pC 1/1-SP31-ADH-GAP

FIG. 23

```
                                                    SOD--►
                                                    MetAlaThrLysAla
                                                    ATGGCTACAAAGGCT
                                                    TACCGATGTTTCCGA

                ValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsnPheGluGlnLysGlu
1383            GTTTGTGTTTTGAAGGGTGACGGCCCAGTTCAAGGTATTATTAACTTCGAGCAGAAGGAA
                CAAACACAAAACTTCCCACTGCCGGGTCAAGTTCCATAATAATTGAAGCTCGTCTTCCTT

                SerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThrGluGlyLeuHisGly
1443            AGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACTGAAGGCCTGCATGGA
                TCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGACTTCCGGACGTACCT

                PheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSerAlaGlyProHisPhe
1503            TTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGTGCAGGTCCTCACTTT
                AAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCACGTCCAGGAGTGAAA

                AsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArgHisValGlyAspLeu
1563            AATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGGCATGTTGGAGACTTG
                TTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCCGTACAACCTCTGAAC

                GlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIleGluAspSerValIle
1623            GGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATTGAAGATTCTGTGATC
                CCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAACTTCTAAGACACTAG

                SerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValValHisGluLysAlaAsp
1683            TCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTCCATGAAAAAGCAGAT
                AGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAGGTACTTTTTCGTCTA

                AspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsnAlaGlySerArgLeu
1743            GACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAACGCTGGAAGTCGTTTG
                CTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTGCGACCTTCAGCAAAC

                                                linker --►          p31 --►
                AlaCysGlyValIleGlyIleAlaGlnAsnSerGlyValGlyAlaMetAlaMetAlaSer
1803            GCTTGTGGTGTAATTGGGATCGCCCAGAATTCAGGTGTTGGAGCCATGGCCATGGCTAGT
                CGAACACCACATTAACCCTAGCGGGTCTTAAGTCCACAACCTCGGTACCGGTACCGATCA

                AspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLysCysGln
1863            GATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAG
                CTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCGACACTATTTACAGTC

                LeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAsp
1923            CTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGAT
                GATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCTTATACCGTTGATCTA
```

FIG. 24A

```
      CysThrHisLeuGluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIle
1983  TGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATA
      ACATGTGTAGATCTTCCTTTTTAATAGGACCATCGTCAAGTACATCGGTCACCTATATAT

      GluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeu
2043  GAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTA
      CTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATAAAAGAGAATTTTAAT

      AlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThr
2103  GCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACT
      CGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCGTTAAAGTGGTCATGA

      ThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsn
2163  ACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAAT
      TGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAACCGTAAGGGATGTTA

      ProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGln
2223  CCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAG
      GGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTCTTTTAATATCCTGTC

      ValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsn
2283  GTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAAT
      CATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGTCATAAGTAGGTGTTA

      PheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArgIleValAspIleIle
2343  TTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATA
      AAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCTTATCATCTGTATTAT

      AlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArg
2403  GCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGG
      CGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTTTAAGTTTTAAAAGCC

      ValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLys
2463  GTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAA
      CAAATAATGTCCCTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGTTTCGAAGAGACCTTT

      GlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysValValProArgArgLys
2523  GGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAA
      CCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCATCACGGTTCTTCTTTT

      AlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArg
2583  GCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGA
      CGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTAACACACCGTTCATCT

      GlnAspGluAspAM
2643  CAGGATGAGGATTAG
      GTCCTACTCCTAATC
```

FIG. 24B

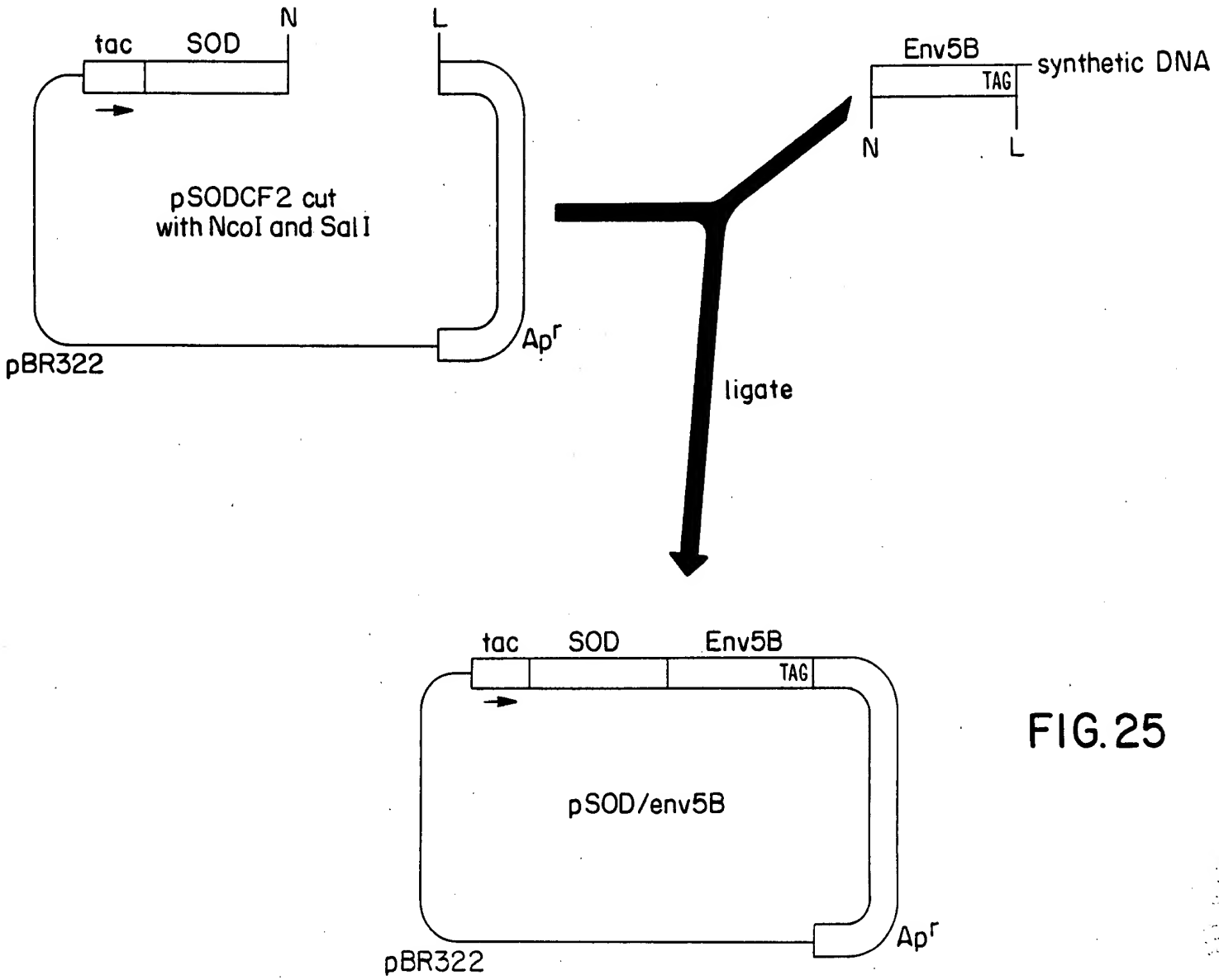


FIG. 25

Sequence of SOD/env-4

```
    SOD -->
    MetAlaThrLysAlaValCysValLeuLysGlyAspGlyProValGlnGlyIleIleAsn
  1 CATGGCGACGAAGGCCGTGTGCGTGCTGAAGGGCGACGGCCCAGTGCAGGGCATCATCAAT
    CGCTGCTTCCGGCACACGCACGACTTCCCGCTGCCGGGTCACGTCCCGTAGTAGTTA

    PheGluGlnLysGluSerAsnGlyProValLysValTrpGlySerIleLysGlyLeuThr
 62 TTCGAGCAGAAGGAAAGTAATGGACCAGTGAAGGTGTGGGGAAGCATTAAAGGACTGACT
    AAGCTCGTCTTCCTTTCATTACCTGGTCACTTCCACACCCCTTCGTAATTTCCTGACTGA

    GluGlyLeuHisGlyPheHisValHisGluPheGlyAspAsnThrAlaGlyCysThrSer
122 GAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATAATACAGCAGGCTGTACCAGT
    CTTCCGGACGTACCTAAGGTACAAGTACTCAAACCTCTATTATGTCGTCCGACATGGTCA

    AlaGlyProHisPheAsnProLeuSerArgLysHisGlyGlyProLysAspGluGluArg
182 GCAGGTCCTCACTTTAATCCTCTATCCAGAAAACACGGTGGGCCAAAGGATGAAGAGAGG
    CGTCCAGGAGTGAAATTAGGAGATAGGTCTTTTGTGCCACCCGGTTTCCTACTTCTCTCC

    HisValGlyAspLeuGlyAsnValThrAlaAspLysAspGlyValAlaAspValSerIle
242 CATGTTGGAGACTTGGGCAATGTGACTGCTGACAAAGATGGTGTGGCCGATGTGTCTATT
    GTACAACCTCTGAACCCGTTACACTGACGACTGTTTCTACCACACCGGCTACACAGATAA

    GluAspSerValIleSerLeuSerGlyAspHisCysIleIleGlyArgThrLeuValVal
302 GAAGATTCTGTGATCTCACTCTCAGGAGACCATTGCATCATTGGCCGCACACTGGTGGTC
    CTTCTAAGACACTAGAGTGAGAGTCCTCTGGTAACGTAGTAACCGGCGTGTGACCACCAG

    HisGluLysAlaAspAspLeuGlyLysGlyGlyAsnGluGluSerThrLysThrGlyAsn
362 CATGAAAAAGCAGATGACTTGGGCAAAGGTGGAAATGAAGAAAGTACAAAGACAGGAAAC
    GTACTTTTTCGTCTACTGAACCCGTTTCCACCTTTACTTCTTTCATGTTTCTGTCCTTTG

                                                    Env4-->
    AlaGlySerArgLeuAlaCysGlyValIleGlhIleAlaMetGluValValIleArgSer
422 GCTGGAAGTCGTTTGGCTTGTGGTGTAATTGGGATCGCCATGGAGGTAGTAATTAGATCT
    CGACCTTCAGCAAACCGAACACCACATTAACCCTAGCGGTACCTCCATCATTAATCTAGA

    AspAsnPheThrAsnAsnAlaLysThrIleIleValGlnLeuAsnGluSerValAlaIle
482 GACAATTTCACGAACAATGCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATT
    CTGTTAAAGTGCTTGTTACGATTTTGGTATTATCATGTCGACTTACTTAGACATCGTTAA

    AsnCysThrArgProAsnAsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAla
542 AACTGTACAAGACCCAACAACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCA
    TTGACATGTTCTGGGTTGTTGTTATGTTCTTTTTCATAGATATATCCTGGTCCCTCTCGT
```

FIG. 26A

```
                PheHisThrThrGlyArgIleIleGlyAspIleArgLysAlaHisCysAsnIleSerArg
602             TTTCATACAACAGGAAGAATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGA
                AAAGTATGTTGTCCTTCTTATTATCCTCTATATTCTTTTCGTGTAACATTGTAATCATCT

                AlaGlnTrpAsnAsnThrLeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsn
662             GCACAATGGAATAACACTTTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAAT
                CGTGTTACCTTATTGTGAAATCTTGTCTATCAATTTTTTAATTCTCTTGTCAAACCCTTA

                AsnLysThrIleValPheAsnGlnSerSerGlyGlyAspProGluIleValMetHisSer
722             AATAAAACAATAGTCTTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGT
                TTATTTTGTTATCAGAAATTAGTTAGGAGTCCTCCCCTGGGTCTTTAACATTACGTGTCA

                PheAsnCysArgGlyGluPhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrp
782             TTTAATTGTAGAGGGGAATTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGG
                AAATTAACATCTCCCCTTAAAAAGATGACATTATGTTGTGTTGACAAATTATTATGTACC

                ArgLeuAsnHisThrGluGlyThrLysGlyAsnAspThrIleIleLeuProCysArgIle
842             AGGTTAAATCACACTGAAGGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATA
                TCCAATTTAGTGTGACTTCCTTGATTTCCTTTACTGTGTTAGTATGAGGGTACATCTTAT

                LysGlnIleIleAsnMetTrpGlnGluValGlyLysAlaMetTyrAlaProProIleGly
902             AAACAAATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGA
                TTTGTTTAATATTTGTACACCGTCCTTCATCCTTTTCGTTACATACGGGGAGGGTAACCT

                GlyGlnIleSerCysSerSerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThr
962             GGACAAATTAGTTGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACA
                CCTGTTTAATCAACAAGTAGTTTATAATGTCCCGACGATAATTGTTCTCTACCACCATGT

                AsnValThrAsnAspThrGluValPheArgProGlyGlyGlyAspMetArgAspAsnTrp
1022            AATGTAACTAATGACACCGAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGG
                TTACATTGATTACTGTGGCTCCAGAAGTCTGGACCTCCTCCTCTATACTCCCTGTTAACC

                ArgSerGluLeuTyrLysTyrLysValIleLysIleGluProLeuGlyIleAlaProThr
1082            AGAAGTGAATTATATAAATATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACC
                TCTTCACTTAATATATTTATATTTCATTATTTTTAACTTGGTAATCCTTATCGTGGGTGG

                LysAlaLysArgArgValValGlnArgGluLysArgOP OP
1142            AAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGATGATGAAGCTTG
                TTCCGTTTCTCTTCTCACCACGTCTCTCTTTTTTCTACTACTTCGAACAGCT
```

FIG. 26B

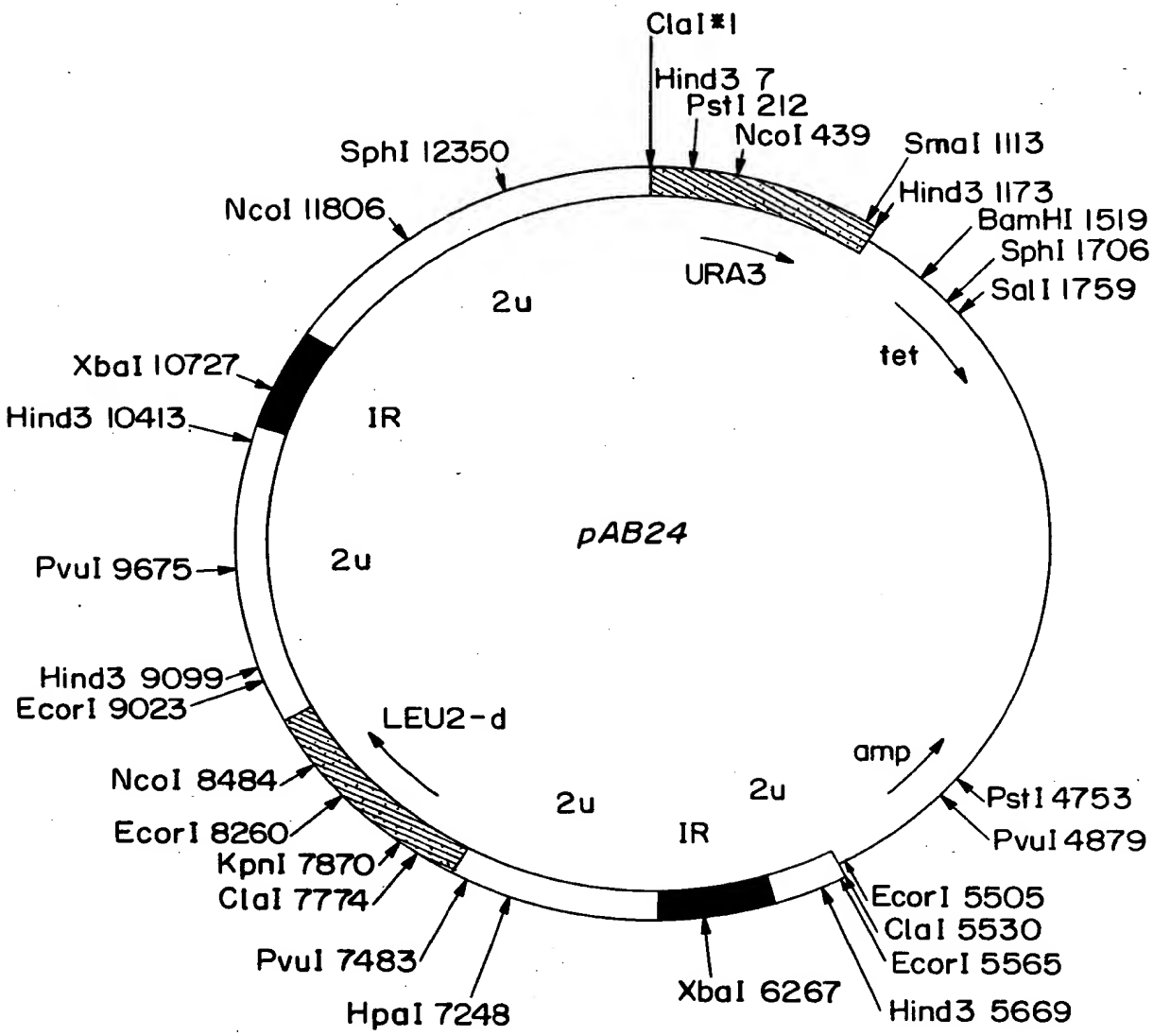

ClaI*1
Hind3 7
PstI 212
NcoI 439
SmaI 1113
Hind3 1173
BamHI 1519
SphI 1706
SalI 1759
SphI 12350
NcoI 11806
URA3
2u
tet
XbaI 10727
Hind3 10413
IR
pAB24
2u
PvuI 9675
Hind3 9099
EcorI 9023
LEU2-d
amp
NcoI 8484
2u
2u
PstI 4753
EcorI 8260
IR
PvuI 4879
KpnI 7870
ClaI 7774
EcorI 5505
ClaI 5530
EcorI 5565
PvuI 7483
XbaI 6267
Hind3 5669
HpaI 7248

FIG. 27

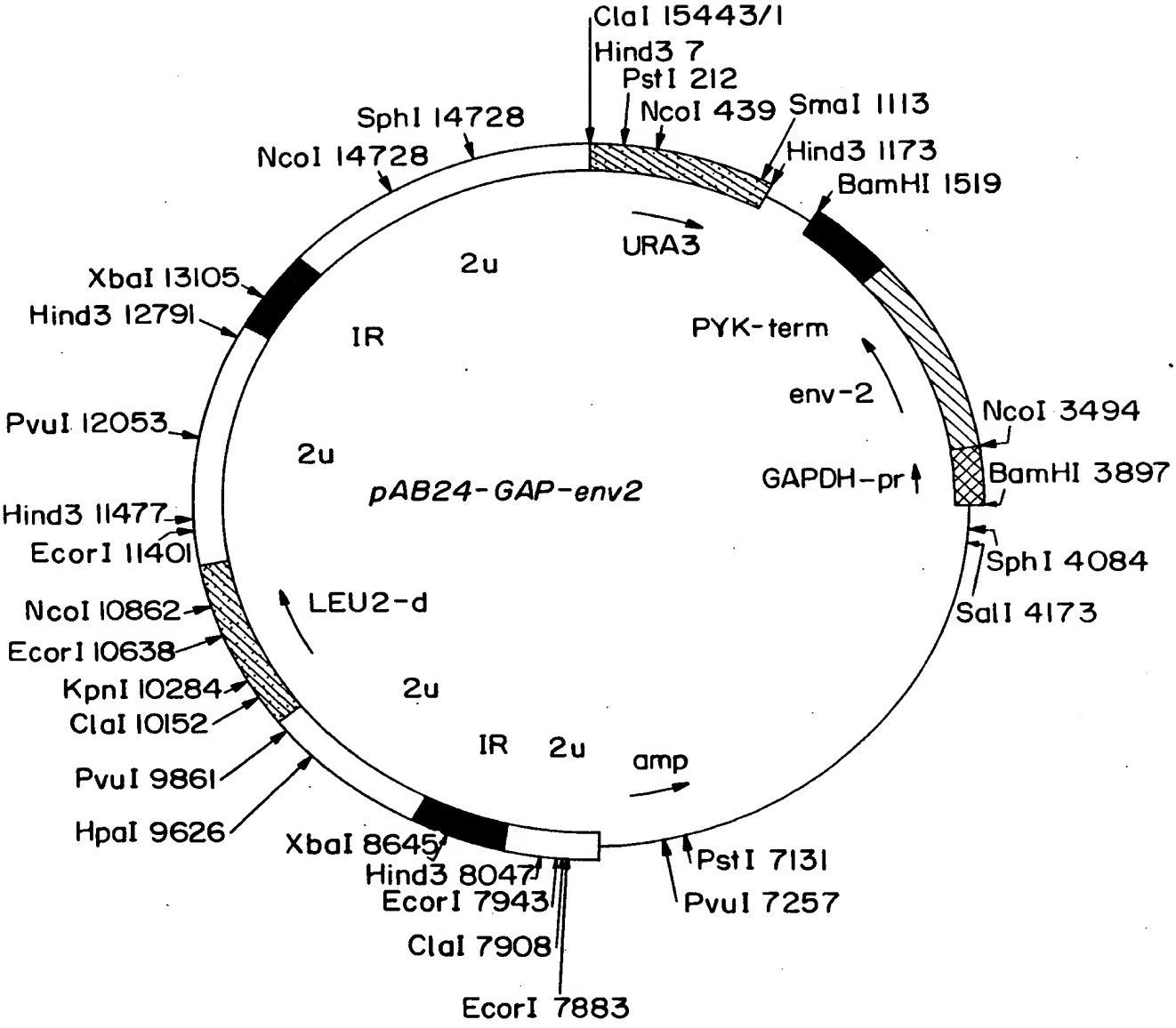

ClaI 15443/1
Hind3 7
PstI 212
NcoI 439
SmaI 1113
Hind3 1173
BamHI 1519
SphI 14728
NcoI 14728
URA3
2u
XbaI 13105
Hind3 12791
IR
PYK-term
env-2
PvuI 12053
NcoI 3494
2u
pAB24-GAP-env2
GAPDH-pr
BamHI 3897
Hind3 11477
EcorI 11401
SphI 4084
LEU2-d
NcoI 10862
SalI 4173
EcorI 10638
KpnI 10284
2u
ClaI 10152
IR
2u
amp
PvuI 9861
HpaI 9626
XbaI 8645
PstI 7131
Hind3 8047
EcorI 7943
PvuI 7257
ClaI 7908
EcorI 7883

FIG. 28

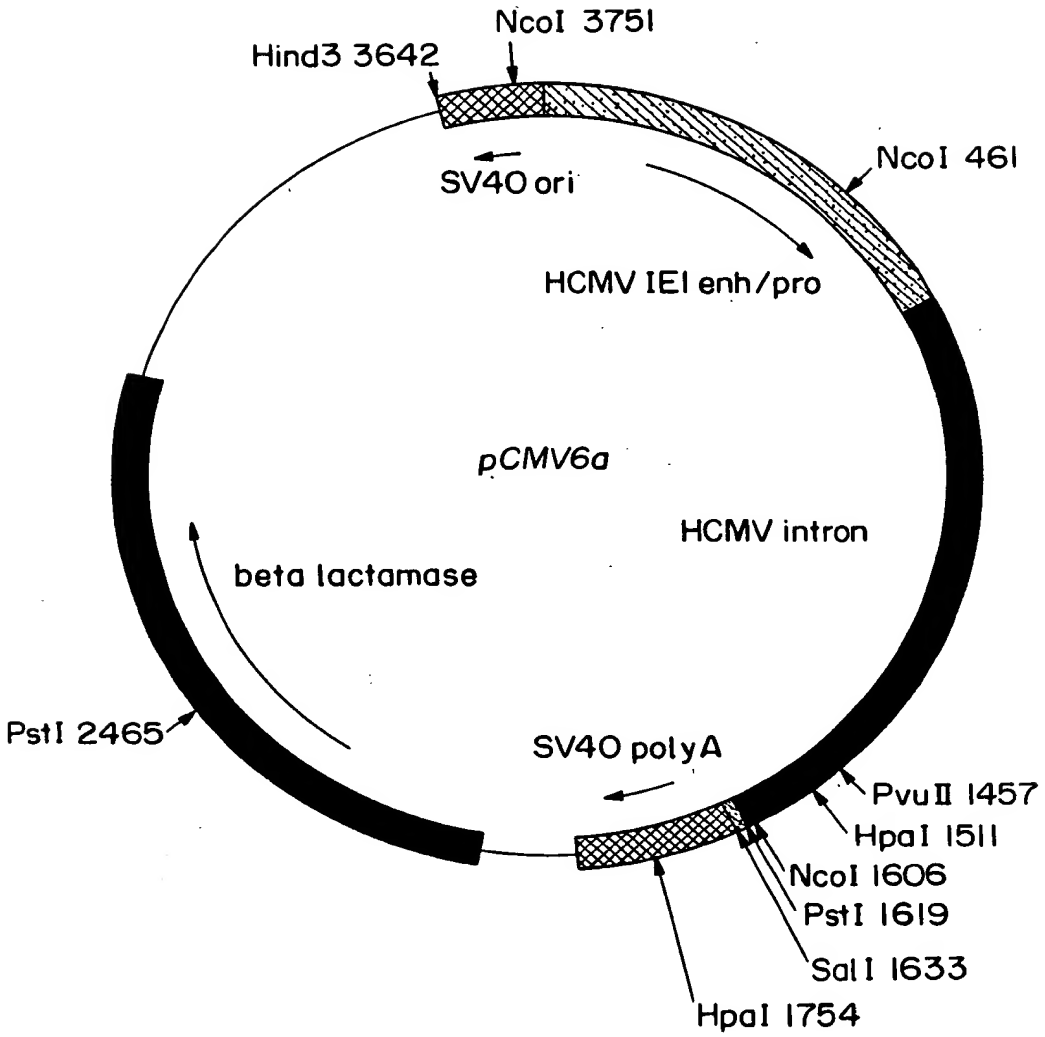

NcoI 3751
Hind3 3642
NcoI 461
SV40 ori
HCMV IEI enh/pro
pCMV6a
HCMV intron
beta lactamase
PstI 2465
SV40 polyA
PvuII 1457
HpaI 1511
NcoI 1606
PstI 1619
SalI 1633
HpaI 1754

FIG. 29

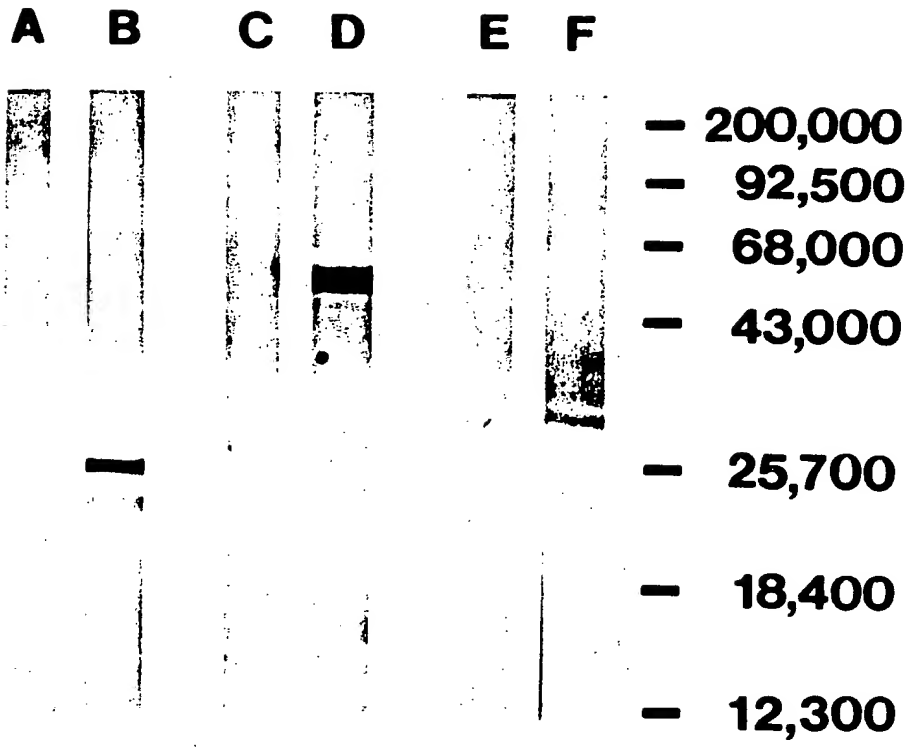
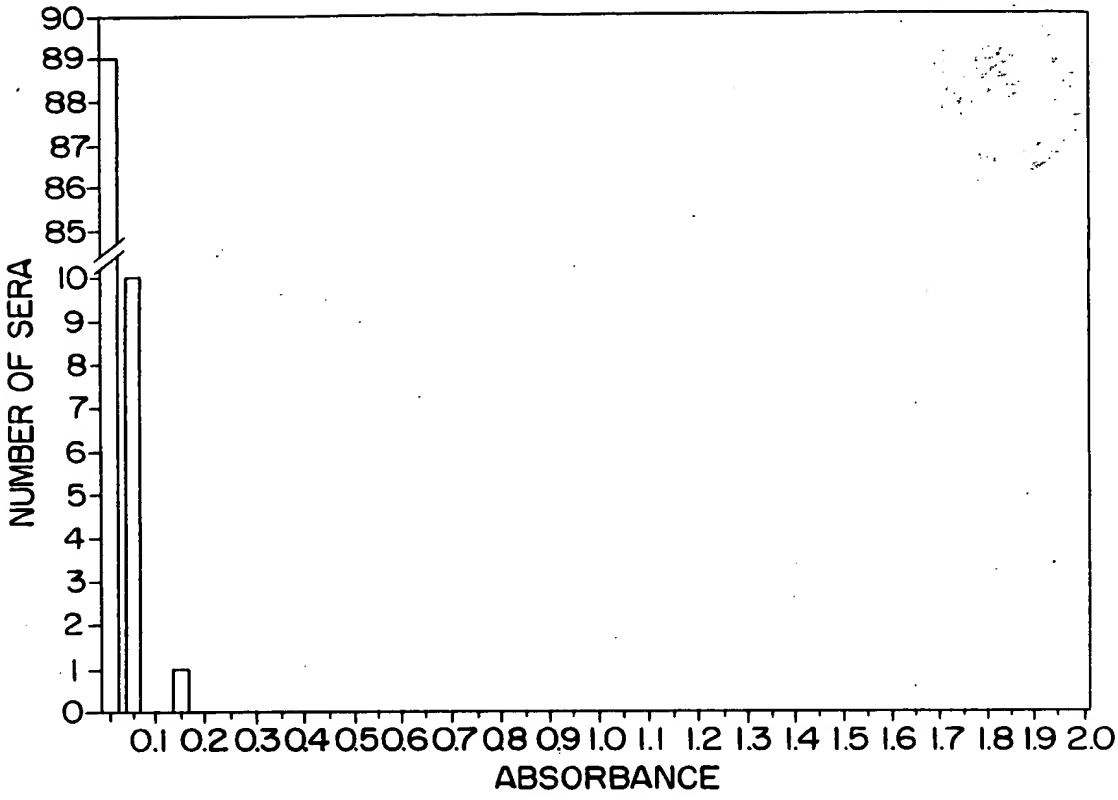


FIG. 30

FIG. 31a



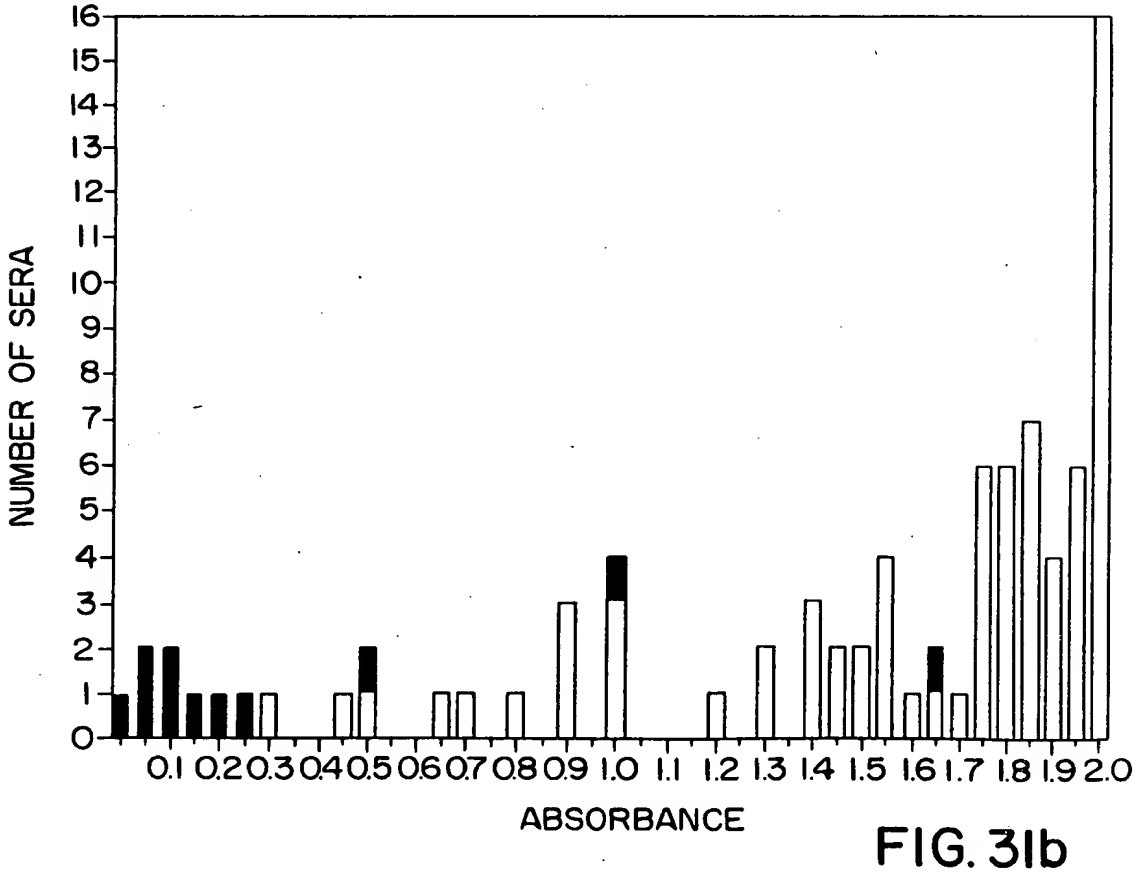


FIG. 31b